I (2)

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: **WO 00/26408**

(43) International Publication Date: 11 May 2000 (11.05.00)

(21) International Application Number: PCT/US99/25431

(22) International Filing Date: 29 October 1999 (29.10.99)

(30) Priority Data:

60/106,308	30 October 1998 (30.10.98)	US
60/136,078	26 May 1999 (26.05.99)	US

(71) Applicant *(for all designated States except US)*: CELLOMICS, INC. [US/US]; 635 William Pitt Way, Pittsburgh, PA 15238 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GUILIANO, Kenneth, A. [US/US]; 351 Hawthorne Road, Pittsburgh, PA 15209 (US). BRIGHT, Gary [US/US]; 586 Sturbridge Drive, Highland Heights, OH 44143 (US). OLSON, Keith [US/US]; 372 S. Highland Avenue, Pittsburgh, PA 15205 (US). BURROUGHS–TENCZA, Sarah [US/US]; 401 Olympia Road, Pittsburgh, PA 15211 (US).

(74) Agent: HALLORAN, Patrick, J.; McDonnell, Boehnen, Hulbert & Berghoff, Suite 3200, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: A SYSTEM FOR CELL–BASED SCREENING

(57) Abstract

The present invention provides systems, methods, screens, reagents and kits for optical system analysis of cells to rapidly determine the distribution, environment, or activity of fluorescently labeled reporter molecules in cells for the purpose of screening large numbers of compounds for those that specifically affect particular biological functions.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	[illegible]	LR	[illegible]	SG	[illegible]		

A SYSTEM FOR CELL-BASED SCREENING

Cross Reference

This application is a continuation-in-part of U.S. Applications for Patent Serial Nos. 60/136,078 filed May 26, 1999 60/106,308 filed October 30, 1998; and 09/398,965 filed September 17, 1999 which is a continuation in part of Serial No. 09/031,271 filed February 27, 1998 which is a continuation in part of U.S. Application S/N 08/810983, filed on February 27, 1997.

Field of The Invention

This invention is in the field of fluorescence-based cell and molecular biochemical assays for drug discovery.

Background of the Invention

Drug discovery, as currently practiced in the art, is a long, multiple step process involving identification of specific disease targets, development of an assay based on a specific target, validation of the assay, optimization and automation of the assay to produce a screen, high throughput screening of compound libraries using the assay to identify "hits", hit validation and hit compound optimization. The output of this process is a lead compound that goes into pre-clinical and, if validated, eventually into clinical trials. In this process, the screening phase is distinct from the assay development phases, and involves testing compound efficacy in living biological systems.

Historically, drug discovery is a slow and costly process, spanning numerous years and consuming hundreds of millions of dollars per drug created. Developments

[illegible] Significant advances in automated DNA sequencing, PCR application, positional cloning, hybridization arrays, and bioinformatics have greatly increased the

number of genes (and gene fragments) encoding potential drug screening targets. However, the basic scheme for drug screening remians the same.

Validation of genomic targets as points for therapeutic intervention using the existing methods and protocols has become a bottleneck in the drug discovery process due to the slow, manual methods employed, such as *in vivo* functional models, functional analysis of recombinant proteins, and stable cell line expression of candidate genes. Primary DNA sequence data acquired through automated sequencing does not permit identification of gene function, but can provide information about common "motifs" and specific gene homology when compared to known sequence databases. Genomic methods such as subtraction hybridization and RADE (rapid amplification of differential expression) can be used to identify genes that are up or down regulated in a disease state model. However, identification and validation still proceed down the same pathway. Some proteomic methods use protein identification (global expression arrays, 2D electrophoresis, combinatorial libraries) in combination with reverse genetics to identify candidate genes of interest. Such putative "disease associated sequences" or DAS isolated as intact cDNA are a great advantage to these methods, but they are identified by the hundreds without providing any information regarding type, activity, and distribution of the encoded protein. Choosing a subset of DAS as drug screening targets is "random", and thus extremely inefficient, without functional data to provide a mechanistic link with disease. It is necessary, therefore, to provide new technologies to rapidly screen DAS to establish biological function, thereby improving target validation and candidate optimization in drug discovery.

There are three major avenues for improving early drug discovery productivity. First, there is a need for tools that provide increased information handling capability. Bioinformatics has blossomed with the rapid development of DNA sequencing systems and the evolution of the genomics database. Genomics is beginning to play a critical role in the identification of potential new targets. Proteomics has become indispensible in relating structure and function of protein targets in order to predict drug interactions.

There is a need for higher throughput tools. Automation is a key to improving productivity as has already been demonstrated in DNA sequencing and high throughput

primary screening. The instant invention provides for automated systems that extract multiple parameter information from cells that meet the need for higher throughput tools. The instant invention also provides for miniaturizing the methods, thereby allowing increased throughput, while decreasing the volumes of reagents and test compounds required in each assay.

Radioactivity has been the dominant read-out in early drug discovery assays. However, the need for more information, higher throughput and miniaturization has caused a shift towards using fluorescence detection. Fluorescence-based reagents can yield more powerful, multiple parameter assays that are higher in throughput and information content and require lower volumes of reagents and test compounds. Fluorescence is also safer and less expensive than radioactivity-based methods.

Screening of cells treated with dyes and fluorescent reagents is well known in the art. There is a considerable body of literature related to genetic engineering of cells to produce fluorescent proteins, such as modified green fluorescent protein (GFP), as a reporter molecule. Some properties of wild-type GFP are disclosed by Morise et al. (*Biochemistry* 13 (1974), p. 2656-2662), and Ward et al. (*Photochem. Photobiol.* 31 (1980), p. 611-615). The GFP of the jellyfish *Aequorea victoria* has an excitation maximum at 395 nm and an emission maximum at 510 nm, and does not require an exogenous factor for fluorescence activity. Uses for GFP disclosed in the literature are widespread and include the study of gene expression and protein localization (Chalfie et al., *Science* 263 (1994), p. 12501-12504)), as a tool for visualizing subcellular organelles (Rizzuto et al., *Curr. Biology* 5 (1995), p. 635-642)), visualization of protein transport along the secretory pathway (Kaether and Gerdes, *FEBS Letters* 369 (1995), p. 267-271)), expression in plant cells (Hu and Cheng, *FEBS Letters* 369 (1995), p. 331-334)) and Drosophila embryos (Davis et al., *Dev. Biology* 170 (1995), p. 726-729)), and as a reporter molecule fused to another protein of interest (U. S. Patent 5,491,084). Similarly, WO96/23898 relates to methods of detecting biologically active substances affecting intracellular processes by utilizing a GFP construct having a

Numerous references are related to GFP proteins in biological systems. For example, WO 96/09598 describes a system for isolating cells of interest utilizing the

expression of a GFP like protein. WO 96/27675 describes the expression of GFP in plants. WO 95/21191 describes modified GFP protein expressed in transformed organisms to detect mutagenesis. U. S. Patents 5,401,629 and 5,436,128 describe assays and compositions for detecting and evaluating the intracellular transduction of an extracellular signal using recombinant cells that express cell surface receptors and contain reporter gene constructs that include transcriptional regulatory elements that are responsive to the activity of cell surface receptors.

Performing a screen on many thousands of compounds requires parallel handling and processing of many compounds and assay component reagents. Standard high throughput screens ("HTS") use mixtures of compounds and biological reagents along with some indicator compound loaded into arrays of wells in standard microtiter plates with 96 or 384 wells. The signal measured from each well, either fluorescence emission, optical density, or radioactivity, integrates the signal from all the material in the well giving an overall population average of all the molecules in the well.

Science Applications International Corporation (SAIC) 130 Fifth Avenue, Seattle, WA. 98109) describes an imaging plate reader. This system uses a CCD camera to image the whole area of a 96 well plate. The image is analyzed to calculate the total fluorescence per well for all the material in the well.

Molecular Devices, Inc. (Sunnyvale, CA) describes a system (FLIPR) which uses low angle laser scanning illumination and a mask to selectively excite fluorescence within approximately 200 microns of the bottoms of the wells in standard 96 well plates in order to reduce background when imaging cell monolayers. This system uses a CCD camera to image the whole area of the plate bottom. Although this system measures signals originating from a cell monolayer at the bottom of the well, the signal measured is averaged over the area of the well and is therefore still considered a measurement of the average response of a population of cells. The image is analyzed to calculate the total fluorescence per well for cell-based assays. Fluid delivery devices have also been incorporated into cell based screening systems, such as the FLIPR

In contrast to high throughput screens, various high-content screens ("HCS") have been developed to address the need for more detailed information about the

temporal-spatial dynamics of cell constituents and processes. High-content screens automate the extraction of multicolor fluorescence information derived from specific fluorescence-based reagents incorporated into cells (Giuliano and Taylor (1995), *Curr. Op. Cell Biol.* 7:4; Giuliano et al. (1995) *Ann. Rev. Biophys. Biomol. Struct.* 24:405). Cells are analyzed using an optical system that can measure spatial, as well as temporal dynamics. (Farkas et al. (1993) *Ann. Rev. Physiol.* 55:785; Giuliano et al. (1990) In *Optical Microscopy for Biology*. B. Herman and K. Jacobson (eds.), pp. 543-557. Wiley-Liss, New York; Hahn et al (1992) *Nature* 359:736; Waggoner et al. (1996) *Hum. Pathol.* 27:494). The concept is to treat each cell as a "well" that has spatial and temporal information on the activities of the labeled constituents.

The types of biochemical and molecular information now accessible through fluorescence-based reagents applied to cells include ion concentrations, membrane potential, specific translocations, enzyme activities, gene expression, as well as the presence, amounts and patterns of metabolites, proteins, lipids, carbohydrates, and nucleic acid sequences (DeBiasio et al., (1996) *Mol. Biol. Cell.* 7:1259;Giuliano et al., (1995) *Ann. Rev. Biophys. Biomol. Struct.* 24:405; Heim and Tsien, (1996) *Curr. Biol.* 6:178).

High-content screens can be performed on either fixed cells, using fluorescently labeled antibodies, biological ligands, and/or nucleic acid hybridization probes, or live cells using multicolor fluorescent indicators and "biosensors." The choice of fixed or live cell screens depends on the specific cell-based assay required.

Fixed cell assays are the simplest, since an array of initially living cells in a microtiter plate format can be treated with various compounds and doses being tested, then the cells can be fixed, labeled with specific reagents, and measured. No environmental control of the cells is required after fixation. Spatial information is acquired, but only at one time point. The availability of thousands of antibodies, ligands and nucleic acid hybridization probes that can be applied to cells makes this an attractive approach for many types of cell-based screens. The fixation and labeling

cells containing the desired reagents can be screened over time, as well as space. Environmental control of the cells (temperature, humidity, and carbon dioxide) is

required during measurement, since the physiological health of the cells must be maintained for multiple fluorescence measurements over time. There is a growing list of fluorescent physiological indicators and "biosensors" that can report changes in biochemical and molecular activities within cells (Giuliano et al., (1995) *Ann. Rev. Biophys. Biomol. Struct.* 24:405; Hahn et al., (1993) In *Fluorescent and Luminescent Probes for Biological Activity*. W.T. Mason, (ed.), pp. 349-359, Academic Press, San Diego).

The availability and use of fluorescence-based reagents has helped to advance the development of both fixed and live cell high-content screens. Advances in instrumentation to automatically extract multicolor, high-content information has recently made it possible to develop HCS into an automated tool. An article by Taylor, et al. (*American Scientist* 80 (1992), p. 322-335) describes many of these methods and their applications. For example, Proffitt et. al. (*Cytometry* 24: 204-213 (1996)) describe a semi-automated fluorescence digital imaging system for quantifying relative cell numbers in situ in a variety of tissue culture plate formats, especially 96-well microtiter plates. The system consists of an epifluorescence inverted microscope with a motorized stage, video camera, image intensifier, and a microcomputer with a PC-Vision digitizer. Turbo Pascal software controls the stage and scans the plate taking multiple images per well. The software calculates total fluorescence per well, provides for daily calibration, and configures easily for a variety of tissue culture plate formats. Thresholding of digital images and reagents which fluoresce only when taken up by living cells are used to reduce background fluorescence without removing excess fluorescent reagent.

Scanning confocal microscope imaging (Go et al., (1997) *Analytical Biochemistry* 247:210-215; Goldman et al., (1995) *Experimental Cell Research* 221:311-319) and multiphoton microscope imaging (Denk et al., (1990) *Science* 248:73; Gratton et al., (1994) *Proc. of the Microscopical Society of America*, pp. 154-155) are also well established methods for acquiring high resolution images of

against the background. For example, it is possible to resolve internal cytoplasmic features of adherent cells from the features on the cell surface. Because scanning

multiphoton imaging requires very short duration pulsed laser systems to achieve the high photon flux required, fluorescence lifetimes can also be measured in these systems (Lakowicz et al., (1992) *Anal. Biochem.* 202:316-330; Gerrittsen et al. (1997), *J. of Fluorescence* 7:11-15)), providing additional capability for different detection modes. Small, reliable and relatively inexpensive laser systems, such as laser diode pumped lasers, are now available to allow multiphoton confocal microscopy to be applied in a fairly routine fashion.

A combination of the biological heterogeneity of cells in populations (Bright, et al., (1989). *J. Cell. Physiol.* 141:410; Giuliano, (1996) *Cell Motil. Cytoskel.* 35:237)) as well as the high spatial and temporal frequency of chemical and molecular information present within cells, makes it impossible to extract high-content information from populations of cells using existing whole microtiter plate readers. No existing high-content screening platform has been designed for multicolor, fluorescence-based screens using cells that are analyzed individually. Similarly, no method is currently available that combines automated fluid delivery to arrays of cells for the purpose of systematically screening compounds for the ability to induce a cellular response that is identified by HCS analysis, especially from cells grown in microtiter plates. Furthermore, no method exists in the art combining high throughput well-by-well measurements to identify "hits" in one assay followed by a second high content cell-by-cell measurement on the same plate of only those wells identified as hits.

The instant invention provides systems, methods, and screens that combine high throughput screening (HTS) and high content screening (HCS) that significantly improve target validation and candidate optimization by combining many cell screening formats with fluorescence-based molecular reagents and computer-based feature extraction, data analysis, and automation, resulting in increased quantity and speed of data collection, shortened cycle times, and, ultimately, faster evaluation of promising drug candidates. The instant invention also provides for miniaturizing the methods,

SUMMARY OF THE INVENTION

In one aspect, the present invention relates to a method for analyzing cells comprising

- providing cells containing fluorescent reporter molecules in an array of locations,
- treating the cells in the array of locations with one or more reagents,
- imaging numerous cells in each location with fluorescence optics,
- converting the optical information into digital data,
- utilizing the digital data to determine the distribution, environment or activity of the fluorescently labeled reporter molecules in the cells and the distribution of the cells, and
- interpreting that information in terms of a positive, negative or null effect of the compound being tested on the biological function

In this embodiment, the method rapidly determines the distribution, environment, or activity of fluorescently labeled reporter molecules in cells for the purpose of screening large numbers of compounds for those that specifically affect particular biological functions. The array of locations may be a microtiter plate or a microchip which is a microplate having cells in an array of locations. In a preferred embodiment, the method includes computerized means for acquiring, processing, displaying and storing the data received. In a preferred embodiment, the method further comprises automated fluid delivery to the arrays of cells. In another preferred embodiment, the information obtained from high throughput measurements on the same plate are used to selectively perform high content screening on only a subset of the cell locations on the plate.

In another aspect of the present invention, a cell screening system is provided that comprises:

- a high magnification fluorescence optical system having a microscope

[illegible] having a means for moving the plate for proper alignment and focusing on the cell arrays;

- a digital camera;
- a light source having optical means for directing excitation light to cell arrays and a means for directing fluorescent light emitted from the cells to the digital camera; and
- a computer means for receiving and processing digital data from the digital camera wherein the computer means includes a digital frame grabber for receiving the images from the camera, a display for user interaction and display of assay results, digital storage media for data storage and archiving, and a means for control, acquisition, processing and display of results.

In a preferred embodiment, the cell screening system further comprises a computer screen operatively associated with the computer for displaying data. In another preferred embodiment, the computer means for receiving and processing digital data from the digital camera stores the data in a bioinformatics data base. In a further preferred embodiment, the cell screening system further comprises a reader that measures a signal from many or all the wells in parallel. In another preferred embodiment, the cell screening system further comprises a mechanical-optical means for changing the magnification of the system, to allow changing modes between high throughput and high content screening. In another preferred embodiment, the cell screening system further comprises a chamber and control system to maintain the temperature, CO_2 concentration and humidity surrounding the plate at levels required to keep cells alive. In a further preferred embodiment, the cell screening system utilizes a confocal scanning illumination and detection system.

In another aspect of the present invention, a machine readable storage medium comprising a program containing a set of instructions for causing a cell screening system to execute procedures for defining the distribution and activity of specific cellular constituents and processes is provided. In a preferred embodiment, the cell screening system comprises a high magnification fluorescence optical system with a

computer means for receiving and processing the digital data from the digital camera. Preferred embodiments of the machine readable storage medium comprise programs

consisting of a set of instructions for causing a cell screening system to execute the procedures set forth in Figures 9, 11, 12, 13, 14 or 15. Another preferred embodiment comprises a program consisting of a set of instructions for causing a cell screening system to execute procedures for detecting the distribution and activity of specific cellular constituents and processes. In most preferred embodiments, the cellular processes include, but are not limited to, nuclear translocation of a protein, cellular hypertrophy, apoptosis, and protease-induced translocation of a protein.

In another preferred embodiment, a variety of automated cell screening methods are provided, including screens to identify compounds that affect transcription factor activity, protein kinase activity, cell morphology, microtubule structure, apoptosis, receptor internalization, and protease-induced translocation of a protein.

In another aspect, the present invention provides recombinant nucleic acids encoding a protease biosensor, comprising:

a. a first nucleic acid sequence that encodes at least one detectable polypeptide signal;

b. a second nucleic acid sequence that encodes at least one protease recognition site, wherein the second nucleic acid sequence is operatively linked to the first nucleic acid sequence that encodes the at least one detectable polypeptide signal; and

c. a third nucleic acid sequence that encodes at least one reactant target sequence, wherein the third nucleic acid sequence is operatively linked to the second nucleic acid sequence that encodes the at least one protease recognition site.

The present invention also provides the recombinant expression vectors capable of expressing the recombinant nucleic acids encoding protease biosensors, as well as genetically modified host cells that are transfected with the expression vectors.

The invention further provides recombinant protease biosensors, comprising

a. a first domain comprising at least one detectable polypeptide signal;
b. a second domain comprising at least one protease recognition site; and

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a diagram of the components of the cell-based scanning system.

Figure 2 shows a schematic of the microscope subassembly.

Figure 3 shows the camera subassembly.

Figure 4 illustrates cell scanning system process.

Figure 5 illustrates a user interface showing major functions to guide the user.

Figure 6 is a block diagram of the two platform architecture of the Dual Mode System for Cell Based Screening in which one platform uses a telescope lens to read all wells of a microtiter plate and a second platform that uses a higher magnification lens to read individual cells in a well.

Figure 7 is a detail of an optical system for a single platform architecture of the Dual Mode System for Cell Based Screening that uses a moveable 'telescope' lens to read all wells of a microtiter plate and a moveable higher magnification lens to read individual cells in a well.

Figure 8 is an illustration of the fluid delivery system for acquiring kinetic data on the Cell Based Screening System.

Figure 9 is a flow chart of processing step for the cell-based scanning system.

Figure 10 A-J illustrates the strategy of the Nuclear Translocation Assay.

Figure 11 is a flow chart defining the processing steps in the Dual Mode System for Cell Based Screening combining high throughput and high content screening of microtiter plates.

Figure 12 is a flow chart defining the processing steps in the High Throughput mode of the System for Cell Based Screening.

Figure 13 is a flow chart defining the processing steps in the High Content mode of the System for Cell Based Screening.

Figure 14 is a flow chart defining the processing steps required for acquiring kinetic data in the High Content mode of the System for Cell Based Screening.

Figure 15 is a flow chart defining the processing steps performed within a well during

Figure 17 is an example of data from a known stimulator of translocation.

Figure 18 illustrates data presentation on a graphical display.

Figure 19 is an illustration of the data from the High Throughput mode of the System for Cell Based Screening, an example of the data passed to the High Content mode, the data acquired in the high content mode, and the results of the analysis of that data.

Figure 20 shows the measurement of a drug-induced cytoplasm to nuclear translocation.

Figure 21 illustrates a graphical user interface of the measurement shown in Figure 20.

Figure 22 illustrates a graphical user interface, with data presentation, of the measurement shown in Fig. 20.

Figure 23 is a graph representing the kinetic data obtained from the measurements depicted in Fig. 20.

Figure 24 details a high-content screen of drug-induced apoptosis.

Figure 25. Graphs depicting changes in morphology upon induction of apoptosis. Staurosporine (A) and paclitaxel (B) induce classic nuclear fragmentation in L929 cells. BHK cells exhibit concentration dependent changes in response to staurosporine (C), but a more classical response to paclitaxel (D). MCF-7 cells exhibit either nuclear condensation (E) or fragmentation (F) in response to staurosporine and paclitaxel, respectively. In all cases, cells were exposed to the compounds for 30 hours.

Figure 26 illustrates the dose response of cells to staurosporine in terms of both nuclear size and nuclear perimeter convolution.

Figure 27. Graphs depicting induction of apoptosis by staurosporine and paclitaxel leading to changes in peri-nuclear f-actin content. (A, B) Both apoptotic stimulators induce dose-dependent increases in f-actin content in L929 cells. (C) In BHK cells, staurosporine induces a dose-dependent increase in f-actin, whereas paclitaxel (D) produces results that are more variable. (E) MCF-7 cells exhibit either a decrease or increase depending on the concentration of staurosporine. (F) Paclitaxel induced changes in f-actin content were highly variable and not significant. Cells were exposed to the compounds for 30 hours.

Figure 28. Graphs depicting mitochondrial changes in response to induction of decrease in membrane potential (E, staurosporine) or an increase in mitochondrial mass (F, paclitaxel) depending on the stimulus. Cells were exposed to the compounds for 30

hours. 28G is a graph showing the simultaneous measurement of staurosporine effects on mitochondrial mass and mitochondrial potential in BHK cells.

Figure 29 shows the nucleic acid and amino acid sequence for various types of protesae biosensor domains. (A) Signal sequences. (B) Protease recognition sites. (C) Product/Reactant target sequences

Figure 30 shows schematically shows some basic organization of domains in the protease biosensors of the invention.

Figure 31 is a schematic diagram of a specific 3-domain protease biosensor.

Figure 32 shows the nucleic acid (SEQ ID NO:1) and amino acid sequence (SEQ ID NO:2) of a specific 3-domain biosensor. The signal domain is in italics, the protease recognition domain is in bold, and the reactant targeting site is underlined.

Figure 33 is a photograph showing the effect of stimulation of apoptosis by cis-platin on BHK cells transfected with an expression vector that expresses the caspase biosensor shown in Figure 32.

Figure 34 shows the nucleic acid (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of a specific 3-domain biosensor. The signal domain is in italics, the protease recognition domain is in bold and the reactant targeting site is unformatted.

Figure 35 shows the nucleic acid (SEQ ID NO:5) and amino acid sequence (SEQ ID NO:6) of a specific 3-domain biosensor. The signal domain is in italics, the protease recognition domain is in bold and the reactant targeting site is unformatted.

Figure 36 shows the nucleic acid (SEQ ID NO:7) and amino acid sequence (SEQ ID NO:8) of a specific 3-domain biosensor. The signal domain is underlined only, the protease recognition domain is in bold and the reactant targeting site is in italics.

Figure 37 shows the nucleic acid (SEQ ID NO:9) and amino acid sequence (SEQ ID NO:10) of a specific 3-domain biosensor. The signal domain is underlined only, the protease recognition domain is in bold and the reactant targeting site is in italics.

Figure 38 is a schematic diagram of a specific 4-domain protease biosensor.

Figure 39 shows the nucleic acid (SEQ ID NO:11) and amino acid sequence (SEQ ID NO:12) of a specific 4-domain biosensor.

Figure 40 shows the nucleic acid (SEQ ID NO:13) and amino acid sequence (SEQ ID NO:14) of a specific 4-domain biosensor.

Figure 41 shows the nucleic acid (SEQ ID NO:15) and amino acid sequence (SEQ ID NO:16) of a specific 4-domain biosensor.

Figure 42 shows the nucleic acid (SEQ ID NO:17) and amino acid sequence (SEQ ID NO:18) of a specific 4-domain biosensor.

Figure 43 shows the nucleic acid (SEQ ID NO:19) and amino acid sequence (SEQ ID NO:20) of a specific 4-domain biosensor.

Figure 44 shows the nucleic acid (SEQ ID NO:21) and amino acid sequence (SEQ ID NO:22) of a specific 4-domain biosensor.

Figure 45 is a schematic diagram of a specific 4-domain protease biosensor, containing a nucleolar localization signal.

Figure 46 shows the nucleic acid (SEQ ID NO:23) and amino acid sequence (SEQ ID NO:24) of a specific 4-domain biosensor.

Figure 47 shows the nucleic acid (SEQ ID NO:25) and amino acid sequence (SEQ ID NO:26) of a specific 4-domain biosensor.

Figure 48 shows the nucleic acid (SEQ ID NO:27) and amino acid sequence (SEQ ID NO:28) of a specific 4-domain biosensor.

Figure 49 shows the nucleic acid (SEQ ID NO:29) and amino acid sequence (SEQ ID NO:30) of a specific 4-domain biosensor.

Figure 50 is a schematic diagram of a specific 5-domain protease biosensor.

Figure 51 shows the nucleic acid (SEQ ID NO:31) and amino acid sequence (SEQ ID NO:32) of a specific 5-domain biosensor.

Figure 52 shows the nucleic acid (SEQ ID NO:33) and amino acid sequence (SEQ ID NO:34) of a specific 5-domain biosensor.

DETAILED DESCRIPTION OF THE INVENTION

All cited patents, patent applications and other references are hereby incorporated by reference in their entirety.

As used herein, the following terms have the specified meaning:

Markers of cellular domains. Luminescent probes that have high affinity for specific cellular constituents including specific organelles or molecules. These probes can either be small luminescent molecules or fluorescently tagged macromolecules used as "labeling reagents", "environmental indicators", or "biosensors."

Labeling reagents. Labeling reagents include, but are not limited to, luminescently labeled macromolecules including fluorescent protein analogs and biosensors, luminescent macromolecular chimeras including those formed with the green fluorescent protein and mutants thereof, luminescently labeled primary or secondary antibodies that react with cellular antigens involved in a physiological response, luminescent stains, dyes, and other small molecules.

Markers of cellular translocations. Luminescently tagged macromolecules or organelles that move from one cell domain to another during some cellular process or physiological response. Translocation markers can either simply report location relative to the markers of cellular domains or they can also be "biosensors" that report some biochemical or molecular activity as well.

Biosensors. Macromolecules consisting of a biological functional domain and a luminescent probe or probes that report the environmental changes that occur either internally or on their surface. A class of luminescently labeled macromolecules designed to sense and report these changes have been termed "fluorescent-protein biosensors". The protein component of the biosensor provides a highly evolved molecular recognition moiety. A fluorescent molecule attached to the protein component in the proximity of an active site transduces environmental changes into fluorescence signals that are detected using a system with an appropriate temporal and spatial resolution such as the cell scanning system of the present invention. Because

activity, these biosensors are essentially reusable.

Disease associated sequences ("DAS"). This term refers to nucleic acid sequences identified by standard techniques, such as primary DNA sequence data, genomic methods such as subtraction hybridization and RADE, and proteomic methods in combination with reverse genetics, as being of drug candidate compounds. The term does not mean that the sequence is only associated with a disease state.

High content screening (HCS) can be used to measure the effects of drugs on complex molecular events such as signal transduction pathways, as well as cell functions including, but not limited to, apoptosis, cell division, cell adhesion, locomotion, exocytosis, and cell-cell communication. Multicolor fluorescence permits multiple targets and cell processes to be assayed in a single screen. Cross-correlation of cellular responses will yield a wealth of information required for target validation and lead optimization.

In one aspect of the present invention, a cell screening system is provided comprising a high magnification fluorescence optical system having a microscope objective, an XY stage adapted for holding a plate with an array of locations for holding cells and having a means for moving the plate to align the locations with the microscope objective and a means for moving the plate in the direction to effect focusing; a digital camera; a light source having optical means for directing excitation light to cells in the array of locations and a means for directing fluorescent light emitted from the cells to the digital camera; and a computer means for receiving and processing digital data from the digital camera wherein the computer means includes: a digital frame grabber for receiving the images from the camera, a display for user interaction and display of assay results, digital storage media for data storage and archiving, and means for control, acquisition, processing and display of results.

Figure 1 is a schematic diagram of a preferred embodiment of the cell scanning system. An inverted fluorescence microscope is used 1, such as a Zeiss Axiovert inverted fluorescence microscope which uses standard objectives with magnification of 1-100x to the camera, and a white light source (e.g., 100W mercury arc lamp or 75W

direction for the microscope objective. A Z axis focus drive 5 moves the objective in the Z direction for focusing. A joystick 6 provides for manual movement of the stage in the XYZ direction. A high resolution digital camera 7 acquires images

from each well or location on the plate. There is a camera power supply 8, an automation controller 9 and a central processing unit 10. The PC 11 provides a display 12 and has associated software. The printer 13 provides for printing of a hard copy record.

Figure 2 is a schematic of one embodiment of the microscope assembly 1 of the invention, showing in more detail the XY stage 3, Z-axis focus drive 5, joystick 6, light source 2, and automation controller 9. Cables to the computer 15 and microscope 16, respectively, are provided. In addition, Figure 2 shows a 96 well microtiter plate 17 which is moved on the XY stage 3 in the XY direction. Light from the light source 2 passes through the PC controlled shutter 18 to a motorized filter wheel 19 with excitation filters 20. The light passes into filter cube 25 which has a dichroic mirror 26 and an emission filter 27. Excitation light reflects off the dichroic mirror to the wells in the microtiter plate 17 and fluorescent light 28 passes through the dichroic mirror 26 and the emission filter 27 and to the digital camera 7.

Figure 3 shows a schematic drawing of a preferred camera assembly. The digital camera 7, which contains an automatic shutter for exposure control and a power supply 31, receives fluorescent light 28 from the microscope assembly. A digital cable 30 transports digital signals to the computer.

The standard optical configurations described above use microscope optics to directly produce an enlarged image of the specimen on the camera sensor in order to capture a high resolution image of the specimen. This optical system is commonly referred to as 'wide field' microscopy. Those skilled in the art of microscopy will recognize that a high resolution image of the specimen can be created by a variety of other optical systems, including, but not limited to, standard scanning confocal detection of a focused point or line of illumination scanned over the specimen (Go et al. 1997, *supra*), and multi-photon scanning confocal microscopy (Denk et al., 1990, *supra*), both of which can form images on a CCD detector or by synchronous digitization of the analog output of a photomultiplier tube.

[illegible] culture, [illegible] take advantage [illegible] particular features [illegible] those cells. [illegible] those skilled in the art of cell culture will recognize that some cell lines are contact inhibited, meaning that they will stop growing when they become surrounded by other cells.

while other cell lines will continue to grow under those conditions and the cells will literally pile up, forming many layers. An example of such a cell line is the HEK 293 (ATCC CRL-1573) line. An optical system that can acquire images of single cell layers in multilayer preparations is required for use with cell lines that tend to form layers. The large depth of field of wide field microscopes produces an image that is a projection through the many layers of cells, making analysis of subcellular spatial distributions extremely difficult in layer-forming cells. Alternatively, the very shallow depth of field that can be achieved on a confocal microscope, (about one micron), allows discrimination of a single cell layer at high resolution, simplifying the determination of the subcellular spatial distribution. Similarly, confocal imaging is preferable when detection modes such as fluorescence lifetime imaging are required.

The output of a standard confocal imaging attachment for a microscope is a digital image that can be converted to the same format as the images produced by the other cell screening system embodiments described above, and can therefore be processed in exactly the same way as those images. The overall control, acquisition and analysis in this embodiment is essentially the same. The optical configuration of the confocal microscope system, is essentially the same as that described above, except for the illuminator and detectors. Illumination and detection systems required for confocal microscopy have been designed as accessories to be attached to standard microscope optical systems such as that of the present invention (Zeiss, Germany). These alternative optical systems therefore can be easily integrated into the system as described above.

Figure 4 illustrates an alternative embodiment of the invention in which cell arrays are in microwells 40 on a microplate 41, described ion co-pending U.S. Application S/N 08/865,341, incorporated by reference herein in its entirety. Typically the microplate is 20 mm by 30 mm as compared to a standard 96 well microtiter plate which is 86 mm by 129 mm. The higher density array of cells on a microplate allows the microplate to be imaged at a low resolution of a few microns per pixel for high throughput modes and at a high resolution of less than one micron per pixel for high resolution modes and to improve the overall throughput of the system.

The microplate chamber 42 serves as a microfluidic delivery system for the addition of compounds to cells. The microplate 41 in the microplate chamber 42 is placed in an XY microplate reader 43. Digital data is processed as described above. The small size of this microplate system increases throughput, minimizes reagent volume and allows control of the distribution and placement of cells for fast and precise cell-based analysis. Processed data can be displayed on a PC screen 11 and made part of a bioinformatics data base 44. This data base not only permits storage and retrieval of data obtained through the methods of this invention, but also permits acquisition and storage of external data relating to cells. Figure 5 is a PC display which illustrates the operation of the software.

In an alternative embodiment, a high throughput system (HTS) is directly coupled with the HCS either on the same platform or on two separate platforms connected electronically (e.g. via a local area network). This embodiment of the invention, referred to as a dual mode optical system, has the advantage of increasing the throughput of a HCS by coupling it with a HTS and thereby requiring slower high resolution data acquisition and analysis only on the small subset of wells that show a response in the coupled HTS.

High throughput 'whole plate' reader systems are well known in the art and are commonly used as a component of an HTS system used to screen large numbers of compounds (Beggs (1997), *J. of Biomolec. Screening* 2:71-78; Macaffrey et al., (1996) *J. Biomolec. Screening* 1:187-190).

In one embodiment of dual mode cell based screening, a two platform architecture in which high throughput acquisition occurs on one platform and high content acquisition occurs on a second platform is provided (Figure 6). Processing occurs on each platform independently, with results passed over a network interface, or a single controller is used to process the data from both platforms.

As illustrated in Figure 6, an exemplified two platform dual mode optical system consists of two light optical instruments [illegible] ...rotter plates or microwell arrays on a microplate, and communicate with each other via an electronic connection 64. The high throughput platform 60 analyzes all the wells in the whole plate either in parallel or rapid serial fashion. Those skilled in the art of

screening will recognize that there are a many such commercially available high throughput reader systems that could be integrated into a dual mode cell based screening system (Topcount (Packard Instruments, Meriden, CT); Spectramax, Lumiskan (Molecular Devices, Sunnyvale, CA); Fluoroscan (Labsystems, Beverly, MA)). The high content platform 65, as described above, scans from well to well and acquires and analyzes high resolution image data collected from individual cells within a well.

The HTS software, residing on the system's computer 62, controls the high throughput instrument, and results are displayed on the monitor 61. The HCS software, residing on it's computer system 67, controls the high content instrument hardware 65, optional devices (e.g. plate loader, environmental chamber, fluid dispenser), analyzes digital image data from the plate, displays results on the monitor 66 and manages data measured in an integrated database. The two systems can also share a single computer, in which case all data would be collected, processed and displayed on that computer, without the need for a local area network to transfer the data. Microtiter plates are transferred from the high throughput system to the high content system 63 either manually or by a robotic plate transfer device, as is well known in the art (Beggs (1997), *supra*; Mcaffrey (1996), *supra*).

In a preferred embodiment, the dual mode optical system utilizes a single platform system (Figure 7). It consists of two separate optical modules, an HCS module 203 and an HTS module 209 that can be independently or collectively moved so that only one at a time is used to collect data from the microtiter plate 201. The microtiter plate 201 is mounted in a motorized X,Y stage so it can be positioned for imaging in either HTS or HCS mode. After collecting and analyzing the HTS image data as described below, the HTS optical module 209 is moved out of the optical path and the HCS optical module 203 is moved into place.

The optical module for HTS 209 consists of a projection lens 214, excitation

lamp system (not illustrated). The fluorescence emission is collected through the

dichroic mirror 210 and emission wavelength filter 211 by a lens 212 which forms an image on the camera 216 with sensor 215.

The optical module for HCS 203 consists of a projection lens 208, excitation wavelength filter 207 and dichroic mirror 204 which are used to illuminate the back aperture of the microscope objective 202, and thereby the field of that objective, from a standard microscope illumination system (not shown). The fluorescence emission is collected by the microscope objective 202, passes through the dichroic mirror 204 and emission wavelength filter 205 and is focused by a tube lens 206 which forms an image on the same camera 216 with sensor 215.

In an alternative embodiment of the present invention, the cell screening system further comprises a fluid delivery device for use with the live cell embodiment of the method of cell screening (see below). Figure 8 exemplifies a fluid delivery device for use with the system of the invention. It consists of a bank of 12 syringe pumps 701 driven by a single motor drive. Each syringe 702 is sized according to the volume to be delivered to each well, typically between 1 and 100 µL. Each syringe is attached via flexible tubing 703 to a similar bank of connectors which accept standard pipette tips 705. The bank of pipette tips are attached to a drive system so they can be lowered and raised relative to the microtiter plate 706 to deliver fluid to each well. The plate is mounted on an X,Y stage, allowing movement relative to the optical system 707 for data collection purposes. This set-up allows one set of pipette tips, or even a single pipette tip, to deliver reagent to all the wells on the plate. The bank of syringe pumps can be used to deliver fluid to 12 wells simultaneously, or to fewer wells by removing some of the tips.

In another aspect, the present invention provides a method for analyzing cells comprising providing an array of locations which contain multiple cells wherein the cells contain one or more fluorescent reporter molecules; scanning multiple cells in each of the locations containing cells to obtain fluorescent signals from the fluorescent reporter molecule in the cells; converting the fluorescent [illegible]

[illegible] reporter molecule within the cells

Cell Arrays

Screening large numbers of compounds for activity with respect to a particular biological function requires preparing arrays of cells for parallel handling of cells and reagents. Standard 96 well microtiter plates which are 86 mm by 129 mm, with 6mm diameter wells on a 9mm pitch, are used for compatibility with current automated loading and robotic handling systems. The microplate is typically 20 mm by 30 mm, with cell locations that are 100-200 microns in dimension on a pitch of about 500 microns. Methods for making microplates are described in U.S. Patent Application Serial No. 08/865,341, incorporated by reference herein in its entirety. Microplates may consist of coplanar layers of materials to which cells adhere, patterned with materials to which cells will not adhere, or etched 3-dimensional surfaces of similarly pattered materials. For the purpose of the following discussion, the terms 'well' and 'microwell' refer to a location in an array of any construction to which cells adhere and within which the cells are imaged. Microplates may also include fluid delivery channels in the spaces between the wells. The smaller format of a microplate increases the overall efficiency of the system by minimizing the quantities of the reagents, storage and handling during preparation and the overall movement required for the scanning operation. In addition, the whole area of the microplate can be imaged more efficiently, allowing a second mode of operation for the microplate reader as described later in this document.

Fluorescence Reporter Molecules

A major component of the new drug discovery paradigm is a continually growing family of fluorescent and luminescent reagents that are used to measure the temporal and spatial distribution, content, and activity of intracellular ions, metabolites, macromolecules, and organelles. Classes of these reagents include labeling reagents that measure the distribution and amount of molecules in living and fixed cells, environmental indicators to report signal transduction events in time and space, and fluorescent protein biosensors to measure target molecular activities within living cell

drug discovery

The method of the present invention is based on the high affinity of fluorescent or luminescent molecules for specific cellular components. The affinity for specific

components is governed by physical forces such as ionic interactions, covalent bonding (which includes chimeric fusion with protein-based chromophores, fluorophores, and lumiphores), as well as hydrophobic interactions, electrical potential, and, in some cases, simple entrapment within a cellular component. The luminescent probes can be small molecules, labeled macromolecules, or genetically engineered proteins, including, but not limited to green fluorescent protein chimeras.

Those skilled in this art will recognize a wide variety of fluorescent reporter molecules that can be used in the present invention, including, but not limited to, fluorescently labeled biomolecules such as proteins, phospholipids and DNA hybridizing probes. Similarly, fluorescent reagents specifically synthesized with particular chemical properties of binding or association have been used as fluorescent reporter molecules (Barak et al., (1997), *J. Biol. Chem.* 272:27497-27500; Southwick et al., (1990), *Cytometry* 11:418-430; Tsien (1989) in *Methods in Cell Biology*, Vol. 29 Taylor and Wang (eds.), pp. 127-156). Fluorescently labeled antibodies are particularly useful reporter molecules due to their high degree of specificity for attaching to a single molecular target in a mixture of molecules as complex as a cell or tissue.

The luminescent probes can be synthesized within the living cell or can be transported into the cell via several non-mechanical modes including diffusion, facilitated or active transport, signal-sequence-mediated transport, and endocytotic or pinocytotic uptake. Mechanical bulk loading methods, which are well known in the art, can also be used to load luminescent probes into living cells (Barber et al. (1996), *Neuroscience Letters* 207:17-20; Bright et al. (1996), *Cytometry* 24:226-233; McNeil (1989) in *Methods in Cell Biology*, Vol. 29, Taylor and Wang (eds.), pp. 153-173). These methods include electroporation and other mechanical methods such as scrape-loading, bead-loading, impact-loading, syringe-loading, hypertonic and hypotonic loading. Additionally, cells can be genetically engineered to express reporter molecules, such as GFP, coupled to a protein of interest as previously described (Chalfie and Prasher U.S. Patent No. 5,491,084, Cubitt et al. 1995

[illegible] the luminescent probes accumulate at their target domain as a result of specific and high affinity interactions with the target domain or other modes of

molecular targeting such as signal-sequence-mediated transport. Fluorescently labeled reporter molecules are useful for determining the location, amount and chemical environment of the reporter. For example, whether the reporter is in a lipophilic membrane environment or in a more aqueous environment can be determined (Giuliano et al. (1995), *Ann. Rev. of Biophysics and Biomolecular Structure* 24:405-434; Giuliano and Taylor (1995), *Methods in Neuroscience* 27:1-16). The pH environment of the reporter can be determined (Bright et al. (1989), *J. Cell Biology* 104:1019-1033; Giuliano et al. (1987), *Anal. Biochem.* 167:362-371; Thomas et al. (1979), *Biochemistry* 18:2210-2218). It can be determined whether a reporter having a chelating group is bound to an ion, such as Ca++, or not (Bright et al. (1989), In *Methods in Cell Biology*, Vol. 30, Taylor and Wang (eds.), pp. 157-192; Shimoura et al. (1988), *J. of Biochemistry* (Tokyo) 251:405-410; Tsien (1989) In *Methods in Cell Biology*, Vol. 30, Taylor and Wang (eds.), pp. 127-156).

Furthermore, certain cell types within an organism may contain components that can be specifically labeled that may not occur in other cell types. For example, epithelial cells often contain polarized membrane components. That is, these cells asymmetrically distribute macromolecules along their plasma membrane. Connective or supporting tissue cells often contain granules in which are trapped molecules specific to that cell type (e.g., heparin, histamine, serotonin, etc.). Most muscular tissue cells contain a sarcoplasmic reticulum, a specialized organelle whose function is to regulate the concentration of calcium ions within the cell cytoplasm. Many nervous tissue cells contain secretory granules and vesicles in which are trapped neurohormones or neurotransmitters. Therefore, fluorescent molecules can be designed to label not only specific components within specific cells, but also specific cells within a population of mixed cell types.

Those skilled in the art will recognize a wide variety of ways to measure fluorescence. For example, some fluorescent reporter molecules exhibit a change in excitation or emission spectra [illegible]

[illegible] exhibit a loss (quenching) or appearance of fluorescence, while some report rotational movements (Giuliano et al. (1995), *Ann. Rev. of Biophysics and Biomol. Structure* 24:405-434; Giuliano et al. (1995), *Methods in Neuroscience* 27:1-16).

Scanning cell arrays

Referring to Figure 9, a preferred embodiment is provided to analyze cells that comprises operator-directed parameters being selected based on the assay being conducted, data acquisition by the cell screening system on the distribution of fluorescent signals within a sample, and interactive data review and analysis. At the start of an automated scan the operator enters information 100 that describes the sample, specifies the filter settings and fluorescent channels to match the biological labels being used and the information sought, and then adjusts the camera settings to match the sample brightness. For flexibility to handle a range of samples, the software allows selection of various parameter settings used to identify nuclei and cytoplasm, and selection of different fluorescent reagents, identification of cells of interest based on morphology or brightness, and cell numbers to be analyzed per well. These parameters are stored in the system's for easy retrieval for each automated run. The system's interactive cell identification mode simplifies the selection of morphological parameter limits such as the range of size, shape, and intensity of cells to be analyzed. The user specifies which wells of the plate the system will scan and how many fields or how many cells to analyze in each well. Depending on the setup mode selected by the user at step 101, the system either automatically pre-focuses the region of the plate to be scanned using an autofocus procedure to "find focus" of the plate 102 or the user interactively pre-focuses 103 the scanning region by selecting three "tag" points which define the rectangular area to be scanned. A least-squares fit "focal plane model" is then calculated from these tag points to estimate the focus of each well during an automated scan. The focus of each well is estimated by interpolating from the focal plane model during a scan.

During an automated scan, the software dynamically displays the scan status, including the number of cells analyzed, the current well being analyzed, images of each independent wavelength as they are acquired, and the result of the screen for each well as it is determined. The plate 4 (Figure 1) is scanned in a serpentine style as the

within each well of a 96-well plate. Those skilled in the programming art will recognize how to adapt software for scanning of other microplate formats such as 24, 48, and 384 well plates. The scan pattern of the entire plate as well as the scan

pattern of fields within each well are programmed. The system adjusts sample focus with an autofocus procedure 104 (Figure 9) through the Z axis focus drive 5, controls filter selection via a motorized filter wheel 19, and acquires and analyzes images of up to four different colors ("channels" or "wavelengths").

The autofocus procedure is called at a user selected frequency, typically for the first field in each well and then once every 4 to 5 fields within each well. The autofocus procedure calculates the starting Z-axis point by interpolating from the pre-calculated plane focal model. Starting a programmable distance above or below this set point, the procedure moves the mechanical Z-axis through a number of different positions, acquires an image at each position, and finds the maximum of a calculated focus score that estimates the contrast of each image. The Z position of the image with the maximum focus score determines the best focus for a particular field. Those skilled in the art will recognize this as a variant of automatic focusing methods as described in Harms et al. in *Cytometry* 5 (1984), 236-243, Groen et al. in *Cytometry* 6 (1985), 81-91, and Firestone et al. in *Cytometry* 12 (1991), 195-206.

For image acquisition, the camera's exposure time is separately adjusted for each dye to ensure a high-quality image from each channel. Software procedures can be called, at the user's option, to correct for registration shifts between wavelengths by accounting for linear (X and Y) shifts between wavelengths before making any further measurements. The electronic shutter 18 is controlled so that sample photo-bleaching is kept to a minimum. Background shading and uneven illumination can be corrected by the software using methods known in the art (Bright et al. (1987), *J. Cell Biol.* 104:1019-1033).

In one channel, images are acquired of a primary marker 105 (Figure 9) (typically cell nuclei counterstained with DAPI or PI fluorescent dyes) which are segmented ("identified") using an adaptive thresholding procedure. The adaptive thresholding procedure 106 is used to dynamically select the threshold of an image for separating cells from the background. The staining of cells with fluorescent dyes can

[illegible] as a result of sample preparation and or the dynamic nature of cells. A global threshold is calculated for the complete image to separate the cells from background and account

for field to field variation. These global adaptive techniques are variants of those described in the art. (Kittler et al. in *Computer Vision, Graphics, and Image Processing* 30 (1985), 125-147, Ridler et al. in *IEEE Trans. Systems, Man, and Cybernetics* (1978), 630-632.)

An alternative adaptive thresholding method utilizes local region thresholding in contrast to global image thresholding. Image analysis of local regions leads to better overall segmentation since staining of cell nuclei (as well as other labeled components) can vary across an image. Using this global/local procedure, a reduced resolution image (reduced in size by a factor of 2 to 4) is first globally segmented (using adaptive thresholding) to find regions of interest in the image. These regions then serve as guides to more fully analyze the same regions at full resolution. A more localized threshold is then calculated (again using adaptive thresholding) for each region of interest.

The output of the segmentation procedure is a binary image wherein the objects are white and the background is black. This binary image, also called a mask in the art, is used to determine if the field contains objects 107. The mask is labeled with a blob labeling method whereby each object (or blob) has a unique number assigned to it. Morphological features, such as area and shape, of the blobs are used to differentiate blobs likely to be cells from those that are considered artifacts. The user pre-sets the morphological selection criteria by either typing in known cell morphological features or by using the interactive training utility. If objects of interest are found in the field, images are acquired for all other active channels 108, otherwise the stage is advanced to the next field 109 in the current well. Each object of interest is located in the image for further analysis 110. The software determines if the object meets the criteria for a valid cell nucleus 111 by measuring its morphological features (size and shape). For each valid cell, the XYZ stage location is recorded, a small image of the cell is stored, and features are measured 112.

The cell scanning method of the present invention can be used to perform many

assay provides for the following measurements:

1. The total fluorescent intensity within the cell nucleus for colors 1-4
2. The area of the cell nucleus for color 1 (the primary marker)
3. The shape of the cell nucleus for color 1 is described by three shape features:
   a) perimeter squared area
   b) box area ratio
   c) height width ratio
4. The average fluorescent intensity within the cell nucleus for colors 1-4 (i.e. #1 divided by #2)
5. The total fluorescent intensity of a ring outside the nucleus (see Figure 10) that represents fluorescence of the cell's cytoplasm (cytoplasmic mask) for colors 2-4
6. The area of the cytoplasmic mask
7. The average fluorescent intensity of the cytoplasmic mask for colors 2-4 (i.e. #5 divided by #6)
8. The ratio of the average fluorescent intensity of the cytoplasmic mask to average fluorescent intensity within the cell nucleus for colors 2-4 (i.e. #7 divided by #4)
9. The difference of the average fluorescent intensity of the cytoplasmic mask and the average fluorescent intensity within the cell nucleus for colors 2-4 (i.e. #7 minus #4)
10. The number of fluorescent domains (also call spots, dots, or grains) within the cell nucleus for colors 2-4

Features 1 through 4 are general features of the different cell screening assays of the invention. These steps are commonly used in a variety of image analysis applications and are well known in art (Russ (1992) *The Image Processing Handbook*, CRC Press Inc.; Gonzales et al. (1987), *Digital Image Processing*. Addison-Wesley Publishing Co. pp. 391-448). Features 5-9 have been developed specifically to provide measurements of a cell's fluorescent molecules within the local cytoplasmic region of the cell and the translocation (i.e. movement) of fluorescent molecules from the cytoplasm to the nucleus. These features (steps 5-9) are used for analyzing cells in microplates for the inhibition of nuclear translocation. For example, inhibition of nuclear translocation of transcription factors provides a novel approach to screening intact cells (detailed examples of other types of screens will be provided below). A specific method measures the [illegible]

[illegible] between these two sub cellular compartments provides a measure of cytoplasm-nuclear translocation (feature 9).

Feature 10 describes a screen used for counting of DNA or RNA probes within the nuclear region in colors 2-4. For example, probes are commercially available for identifying chromosome-specific DNA sequences (Life Technologies, Gaithersburg, MD; Genosys, Woodlands, TX; Biotechnologies, Inc., Richmond, CA; Bio 101, Inc., Vista, CA) Cells are three-dimensional in nature and when examined at a high magnification under a microscope one probe may be in-focus while another may be completely out-of-focus. The cell screening method of the present invention provides for detecting three-dimensional probes in nuclei by acquiring images from multiple focal planes. The software moves the Z-axis motor drive 5 (Figure 1) in small steps where the step distance is user selected to account for a wide range of different nuclear diameters. At each of the focal steps, an image is acquired. The maximum gray-level intensity from each pixel in each image is found and stored in a resulting maximum projection image. The maximum projection image is then used to count the probes. The above method works well in counting probes that are not stacked directly above or below another one. To account for probes stacked on top of each other in the Z-direction, users can select an option to analyze probes in each of the focal planes acquired. In this mode, the scanning system performs the maximum plane projection method as discussed above, detects probe regions of interest in this image, then further analyzes these regions in all the focal plane images.

After measuring cell features 112 (Figure 9), the system checks if there are any unprocessed objects in the current field 113. If there are any unprocessed objects, it locates the next object 110 and determines whether it meets the criteria for a valid cell nucleus 111, and measures its features. Once all the objects in the current field are processed, the system determines whether analysis of the current plate is complete 114; if not, it determines the need to find more cells in the current well 115. If the need exists, the system advances the XYZ stage to the next field within the current well 109 or advances the stage to the next well 116 of the plate.

After a plate scan is complete, images and data can be reviewed with the

[illegible] database [illegible] after review [illegible]

interfacing with a network information management system. Data can also be exported to other third-party statistical packages to tabulate results and generate other reports.

Users can review the images alone of every cell analyzed by the system with an interactive image review procedure 117. The user can review data on a cell-by-cell basis using a combination of interactive graphs, a data spreadsheet of measured features, and images of all the fluorescence channels of a cell of interest with the interactive cell-by-cell data review procedure 118. Graphical plotting capabilities are provided in which data can be analyzed via interactive graphs such as histograms and scatter plots. Users can review summary data that are accumulated and summarized for all cells within each well of a plate with an interactive well-by-well data review procedure 119. Hard copies of graphs and images can be printed on a wide range of standard printers.

As a final phase of a complete scan, reports can be generated on one or more statistics of the measured features. Users can generate a graphical report of data summarized on a well-by-well basis for the scanned region of the plate using an interactive report generation procedure 120. This report includes a summary of the statistics by well in tabular and graphical format and identification information on the sample. The report window allows the operator to enter comments about the scan for later retrieval. Multiple reports can be generated on many statistics and be printed with the touch of one button. Reports can be previewed for placement and data before being printed.

The above-recited embodiment of the method operates in a single high resolution mode referred to as the high content screening (HCS) mode. The HCS mode provides sufficient spatial resolution within a well (on the order of 1 μm) to define the distribution of material within the well, as well as within individual cells in the well. The high degree of information content accessible in that mode, comes at the expense of speed and complexity of the required signal processing.

In an alternative embodiment, a high throughput system (HTS) is directly coupled with the HCS either on the same platform or on two separate platforms connected electronically (e.g. [illegible]

throughput of an HCS by coupling it with an HTS and thereby requiring slower high resolution data acquisition and analysis only on the small subset of wells that show a response in the coupled HTS.

High throughput 'whole plate' reader systems are well known in the art and are commonly used as a component of an HTS system used to screen large numbers of compounds (Beggs et al. (1997), *supra*; McCaffrey et al. (1996), *supra*). The HTS of the present invention is carried out on the microtiter plate or microwell array by reading many or all wells in the plate simultaneously with sufficient resolution to make determinations on a well-by-well basis. That is, calculations are made by averaging the total signal output of many or all the cells or the bulk of the material in each well. Wells that exhibit some defined response in the HTS (the 'hits') are flagged by the system. Then on the same microtiter plate or microwell array, each well identified as a hit is measured via HCS as described above. Thus, the dual mode process involves:

1. Rapidly measuring numerous wells of a microtiter plate or microwell array,
2. Interpreting the data to determine the overall activity of fluorescently labeled reporter molecules in the cells on a well-by-well basis to identify "hits" (wells that exhibit a defined response),
3. Imaging numerous cells in each "hit" well, and
4. Interpreting the digital image data to determine the distribution, environment or activity of the fluorescently labeled reporter molecules in the individual cells (i.e. intracellular measurements) and the distribution of the cells to test for specific biological functions

In a preferred embodiment of dual mode processing (Figure 11), at the start of a run 301, the operator enters information 302 that describes the plate and its contents, specifies the filter settings and fluorescent channels to match the biological labels being used, the information sought and the camera settings to match the sample brightness. These parameters are stored in the system's database for easy retrieval for each automated run. The microtiter plate or microwell array is loaded into the cell screening system 303 either manually or automatically by controlling a robotic loading device. An optional environmental chamber 304 is controlled by the system to maintain the temperature, humidity and CO_2 levels in the air surrounding live cells in the microtiter plate or microwell array. An optional fluid delivery device 305 (see Figure 8) is

... to dispense processing agents ... are performed ... the microtiter plate or microwell array by acquiring and analyzing the signal from each of the wells in the

plate. The processing performed in high throughput mode 307 is illustrated in Figure 12 and described below. Wells that exhibit some selected intensity response in this high throughput mode ("hits") are identified by the system. The system performs a conditional operation 308 that tests for hits. If hits are found, those specific hit wells are further analyzed in high content (micro level) mode 309. The processing performed in high content mode 312 is illustrated in Figure 13. The system then updates 310 the informatics database 311 with results of the measurements on the plate. If there are more plates to be analyzed 313 the system loads the next plate 303; otherwise the analysis of the plates terminates 314.

The following discussion describes the high throughput mode illustrated in Figure 12. The preferred embodiment of the system, the single platform dual mode screening system, will be described. Those skilled in the art will recognize that operationally the dual platform system simply involves moving the plate between two optical systems rather than moving the optics. Once the system has been set up and the plate loaded, the system begins the HTS acquisition and analysis 401. The HTS optical module is selected by controlling a motorized optical positioning device 402 on the dual mode system. In one fluorescence channel, data from a primary marker on the plate is acquired 403 and wells are isolated from the plate background using a masking procedure 404. Images are also acquired in other fluorescence channels being used 405. The region in each image corresponding to each well 406 is measured 407. A feature calculated from the measurements for a particular well is compared with a predefined threshold or intensity response 408, and based on the result the well is either flagged as a "hit" 409 or not. The locations of the wells flagged as hits are recorded for subsequent high content mode processing. If there are wells remaining to be processed 410 the program loops back 406 until all the wells have been processed 411 and the system exits high throughput mode.

Following HTS analysis, the system starts the high content mode processing

throughput mode, the XY stage location of the well is retrieved from memory or disk and the stage is then moved to the selected stage location 503. The autofocus procedure

504 is called for the first field in each hit well and then once every 5 to 8 fields within each well. In one channel, images are acquired of the primary marker 505 (typically cell nuclei counterstained with DAPI, Hoechst or PI fluorescent dye). The images are then segmented (separated into regions of nuclei and non-nuclei) using an adaptive thresholding procedure 506. The output of the segmentation procedure is a binary mask wherein the objects are white and the background is black. This binary image, also called a mask in the art, is used to determine if the field contains objects 507. The mask is labeled with a blob labeling method whereby each object (or blob) has a unique number assigned to it. If objects are found in the field, images are acquired for all other active channels 508, otherwise the stage is advanced to the next field 514 in the current well. Each object is located in the image for further analysis 509. Morphological features, such as area and shape of the objects, are used to select objects likely to be cell nuclei 510, and discard (do no further processing on) those that are considered artifacts. For each valid cell nucleus, the XYZ stage location is recorded, a small image of the cell is stored, and assay specific features are measured 511. The system then performs multiple tests on the cells by applying several analytical methods to measure features at each of several wavelengths. After measuring the cell features, the systems checks if there are any unprocessed objects in the current field 512. If there are any unprocessed objects, it locates the next object 509 and determines whether it meets the criteria for a valid cell nucleus 510, and measures its features. After processing all the objects in the current field, the system deteremines whether it needs to find more cells or fields in the current well 513. If it needs to find more cells or fields in the current well it advances the XYZ stage to the next field within the current well 515. Otherwise, the system checks whether it has any remaining hit wells to measure 515. If so, it advances to the next hit well 503 and proceeds through another cycle of acquisition and analysis, otherwise the HCS mode is finished 516.

In an alternative embodiment of the present invention, a method of kinetic live cell screening is provided. The previously described embodiments [illegible] chemical fixation. [illegible] such, these embodiments have limited utility for implementing kinetic based screens, due to the sequential nature of the image acquisition, and the amount of time required to read all the wells on a plate. For

example, since a plate can require 30 – 60 minutes to read through all the wells, only very slow kinetic processes can be measured by simply preparing a plate of live cells and then reading through all the wells more than once. Faster kinetic processes can be measured by taking multiple readings of each well before proceeding to the next well, but the elapsed time between the first and last well would be too long, and fast kinetic processes would likely be complete before reaching the last well.

The kinetic live cell extension of the invention enables the design and use of screens in which a biological process is characterized by its kinetics instead of, or in addition to, its spatial characteristics. In many cases, a response in live cells can be measured by adding a reagent to a specific well and making multiple measurements on that well with the appropriate timing. This dynamic live cell embodiment of the invention therefore includes apparatus for fluid delivery to individual wells of the system in order to deliver reagents to each well at a specific time in advance of reading the well. This embodiment thereby allows kinetic measurements to be made with temporal resolution of seconds to minutes on each well of the plate. To improve the overall efficiency of the dynamic live cell system, the acquisition control program is modified to allow repetitive data collection from sub-regions of the plate, allowing the system to read other wells between the time points required for an individual well.

Figure 8 describes an example of a fluid delivery device for use with the live cell embodiment of the invention and is described above. This set-up allows one set of pipette tips 705, or even a single pipette tip, to deliver reagent to all the wells on the plate. The bank of syringe pumps 701 can be used to deliver fluid to 12 wells simultaneously, or to fewer wells by removing some of the tips 705. The temporal resolution of the system can therefore be adjusted, without sacrificing data collection efficiency, by changing the number of tips and the scan pattern as follows. Typically, the data collection and analysis from a single well takes about 5 seconds. Moving from well to well and focusing in a well requires about 5 seconds, so the overall cycle time

made with about 5 seconds temporal resolution. If 6 pipette tips are used to deliver fluids to 6 wells simultaneously, and the system repetitively scans all 6 wells, each scan

will require 60 seconds, thereby establishing the temporal resolution. For slower processes which only require data collection every 8 minutes, fluids can be delivered to one half of the plate, by moving the plate during the fluid delivery phase, and then repetitively scanning that half of the plate. Therefore, by adjusting the size of the sub-region being scanned on the plate, the temporal resolution can be adjusted without having to insert wait times between acquisitions. Because the system is continuously scanning and acquiring data, the overall time to collect a kinetic data set from the plate is then simply the time to perform a single scan of the plate, multiplied by the number of time points required. Typically, 1 time point before addition of compounds and 2 or 3 time points following addition should be sufficient for screening purposes.

Figure 14 shows the acquisition sequence used for kinetic analysis. The start of processing 801 is configuration of the system, much of which is identical to the standard HCS configuration. In addition, the operator must enter information specific to the kinetic analysis being performed 802, such as the sub-region size, the number of time points required, and the required time increment. A sub-region is a group of wells that will be scanned repetitively in order to accumulate kinetic data. The size of the sub-region is adjusted so that the system can scan a whole sub-region once during a single time increment, thus minimizing wait times. The optimum sub-region size is calculated from the setup parameters, and adjusted if necessary by the operator. The system then moves the plate to the first sub-region 803, and to the first well in that sub-region 804 to acquire the prestimulation (time = 0) time points. The acquisition sequence performed in each well is exactly the same as that required for the specific HCS being run in kinetic mode. Figure 15 details a flow chart for that processing. All of the steps between the start 901 and the return 902 are identical to those described as steps 504 – 514 in Figure 13.

After processing each well in a sub-region, the system checks to see if all the wells in the sub-region have been processed 806 (Figure 14), and cycles through all the wells until the sub-region is [illegible]

[illegible] region size may require multiple additions for sub-regions which span several rows on the plate, with the system moving the plate on the X,Y stage between

additions. Once the fluids have been added, the system moves to the first well in the sub-region 808 to begin acquisition of time points. The data is acquired from each well 809 and as before the system cycles through all the wells in the sub-region 810. After each pass through the sub-region, the system checks whether all the time points have been collected 811 and if not, pauses 813 if necessary 812 to stay synchronized with the requested time increment. Otherwise, the system checks for additional sub-regions on the plate 814 and either moves to the next sub-region 803 or finishes 815. Thus, the kinetic analysis mode comprises operator identification of sub-regions of the microtiter plate or microwells to be screened, based on the kinetic response to be investigated, with data acquisitions within a sub-region prior to data acquisition in subsequent sub-regions.

Specific Screens

In another aspect of the present invention, cell screening methods and machine readable storage medium comprising a program containing a set of instructions for causing a cell screening system to execute procedures for defining the distribution and activity of specific cellular constituents and processes is provided. In a preferred embodiment, the cell screening system comprises a high magnification fluorescence optical system with a stage adapted for holding cells and a means for moving the stage, a digital camera, a light source for receiving and processing the digital data from the digital camera, and a computer means for receiving and processing the digital data from the digital camera. This aspect of the invention comprises programs that instruct the cell screening system to define the distribution and activity of specific cellular constituents and processes, using the luminescent probes, the optical imaging system, and the pattern recognition software of the invention. Preferred embodiments of the machine readable storage medium comprise programs consisting of a set of instructions for causing a cell screening system to execute the procedures set forth in Figures 9, 11, 12, 13, 14 or 15. Another preferred embodiment comprises a program consisting of a set of instructions for causing a cell screening system

In preferred embodiments, the cellular processes include, but are not limited to,

nuclear translocation of a protein, cellular morphology, apoptosis, receptor internalization, and protease-induced translocation of a protein.

In a preferred embodiment, the cell screening methods are used to identify compounds that modify the various cellular processes. The cells can be contacted with a test compound, and the effect of the test compound on a particular cellular process can be analyzed. Alternatively, the cells can be contacted with a test compound and a known agent that modifies the particular cellular process, to determine whether the test compound can inhibit or enhance the effect of the known agent. Thus, the methods can be used to identify test compounds that increase or decrease a particular cellular response, as well as to identify test compounds that affects the ability of other agents to increase or decrease a particular cellular response.

In another preferred embodiment, the locations containing cells are analyzed using the above methods at low resolution in a high throughput mode, and only a subset of the locations containing cells are analyzed in a high content mode to obtain luminescent signals from the luminescently labeled reporter molecules in subcellular compartments of the cells being analyzed.

The following examples are intended for purposes of illustration only and should not be construed to limit the scope of the invention, as defined in the claims appended hereto.

The various chemical compounds, reagents, dyes, and antibodies that are referred to in the following Examples are commercially available from such sources as Sigma Chemical (St. Louis, MO), Molecular Probes (Eugene, OR), Aldrich Chemical Company (Milwaukee, WI), Accurate Chemical Company (Westbury, NY), Jackson Immunolabs, and Clontech (Palo Alto, CA).

Example 1 **Cytoplasm to Nucleus Translocation Screening:**

a. Transcription Factors

[illegible] can initiate transcription of a particular gene or genes. This change in transcription factor distribution is the basis of a screen for the cell-based screening system to detect

compounds that inhibit or induce transcription of a particular gene or group of genes. A general description of the screen is given followed by a specific example.

The distribution of the transcription factor is determined by labeling the nuclei with a DNA specific fluorophore like Hoechst 33423 and the transcription factor with a specific fluorescent antibody. After autofocusing on the Hoechst labeled nuclei, an image of the nuclei is acquired in the cell-based screening system and used to create a mask by one of several optional thresholding methods, as described *supra*. The morphological descriptors of the regions defined by the mask are compared with the user defined parameters and valid nuclear masks are identified and used with the following method to extract transcription factor distributions. Each valid nuclear mask is eroded to define a slightly smaller nuclear region. The original nuclear mask is then dilated in two steps to define a ring shaped region around the nucleus, which represents a cytoplasmic region. The average antibody fluorescence in each of these two regions is determined, and the difference between these averages is defined as the NucCyt Difference. Two examples of determining nuclear translocation are discussed below and illustrated in **Figure 10A-J. Figure 10A** illustrates an unstimulated cell with its nucleus 200 labeled with a blue fluorophore and a transcription factor in the cytoplasm 201 labeled with a green fluorophore. **Figure 10B** illustrates the nuclear mask 202 derived by the cell-based screening system. **Figure 10C** illustrates the cytoplasm 203 of the unstimulated cell imaged at a green wavelength. **Figure 10D** illustrates the nuclear mask 202 is eroded (reduced) once to define a nuclear sampling region 204 with minimal cytoplasmic distribution. The nucleus boundary 202 is dilated (expanded) several times to form a ring that is 2-3 pixels wide that is used to define the cytoplasmic sampling region 205 for the same cell. Figure 10E further illustrates a side view which shows the nuclear sampling region 204 and the cytoplasmic sampling region 205. Using these two sampling regions, data on nuclear translocation can be automatically analyzed by the cell-based screening system on a cell by cell basis. **Figure 10F-J** illustrates the strategy for determining nuclear translocation in

[illegible] and a transcription factor in the cytoplasm 201 labeled with a green fluorophore. The nuclear mask 208 in **Figure 10G** is derived by the cell based screening system. **Figure 10H** illustrates the cytoplasm 209 of a stimulated cell imaged

at a green wavelength. **Figure 10I** illustrates the nuclear sampling region 211 and cytoplasmic sampling region 212 of the stimulated cell. **Figure 10J** further illustrates a side view which shows the nuclear sampling region 211 and the cytoplasmic sampling region 212.

A specific application of this method has been used to validate this method as a screen. A human cell line was plated in 96 well microtiter plates. Some rows of wells were titrated with IL-1, a known inducer of the NF-KB transcription factor. The cells were then fixed and stained by standard methods with a fluorescein labeled antibody to the transcription factor, and Hoechst 33423. The cell-based screening system was used to acquire and analyze images from this plate and the NucCyt Difference was found to be strongly correlated with the amount of agonist added to the wells as illustrated in Figure 16. In a second experiment, an antagonist to the receptor for IL-1, IL-1RA was titrated in the presence of IL-1α, progressively inhibiting the translocation induced by IL-1α. The NucCyt Difference was found to strongly correlate with this inhibition of translocation, as illustrated in **Figure 17**.

Additional experiments have shown that the NucCyt Difference, as well as the NucCyt ratio, gives consistent results over a wide range of cell densities and reagent concentrations, and can therefore be routinely used to screen compound libraries for specific nuclear translocation activity. Furthermore, the same method can be used with antibodies to other transcription factors, or GFP-transcription factor chimeras, or fluorescently labeled transcription factors introduced into living or fixed cells, to screen for effects on the regulation of transcription factor activity.

Figure 18 is a representative display on a PC screen of data which was obtained in accordance with Example 1. Graph 1 180 plots the difference between the average antibody fluorescence in the nuclear sampling region and cytoplasmic sampling region, NucCyt Difference verses Well #. Graph 2 181 plots the average fluorescence of the antibody in the nuclear sampling region, NP1 average, versus the Well #. Graph 3 182 plots the average antibody fluorescence in the cytoplasmic sampling region, NP1

[illegible] Figure 19 shows a screen display of the nuclear image 184 and the fluorescent antibody image 185 for cell #26.

NucCyt Difference referred to in graph 1 180 of **Figure 18** is the difference between the average cytoplasmic probe (fluorescent reporter molecule) intensity and the average nuclear probe (fluorescent reporter molecule) intensity. NP1 average referred to in graph 2 181 of **Figure 18** is the average of cytoplasmic probe (fluorescent reporter molecule) intensity within the nuclear sampling region. L1P1 average referred to in graph 3 182 of **Figure 18** is the average probe (fluorescent reporter molecule) intensity within the cytoplasmic sampling region.

It will be understood by one of skill in the art that this aspect of the invention can be performed using other transcription factors that translocate from the cytoplasm to the nucleus upon activation. In another specific example, activation of the c-fos transcription factor was assessed by defining its spatial position within cells. Activated c-fos is found only within the nucleus, while inactivated c-fos resides within the cytoplasm. 3T3 cells were plated at 5000-10000 cells per well in a Polyfiltronics 96-well plate. The cells were allowed to attach and grow overnight. The cells were rinsed twice with 100 µl serum-free medium, incubated for 24-30 hours in serum-free MEM culture medium, and then stimulated with platelet derived growth factor (PDGF-BB) (Sigma Chemical Co., St. Louis, MO) diluted directly into serum free medium at concentrations ranging from 1-50 ng/ml for an average time of 20 minutes.

Following stimulation, cells were fixed for 20 minutes in 3.7% formaldehyde solution in 1X Hanks buffered saline solution (HBSS). After fixation, the cells were washed with HBSS to remove residual fixative, permeabilized for 90 seconds with 0.5% Triton X-100 solution in HBSS, and washed twice with HBSS to remove residual detergent. The cells were then blocked for 15 minutes with a 0.1% solution of BSA in HBSS, and further washed with HBSS prior to addition of diluted primary antibody solution.

c-Fos rabbit polyclonal antibody (Calbiochem, PC05) was diluted 1:50 in HBSS, and 50 µl of the dilution was applied to each well. Cells were incubated in the

[illegible] temperature [illegible] well was contained with goat anti-rabbit secondary antibody conjugated to ALEXA™ 488 (Molecular Probes), diluted 1:500 from a 100 µg/ml stock in HBSS. Hoechst DNA dye (Molecular Probes) was then added at a

1:1000 dilution of the manufacturer's stock solution (10 mg/ml). The cells were then washed with HBSS, and the plate was sealed prior to analysis with the cell screening system of the invention. The data from these experiments demonstrated that the methods of the invention could be used to measure transcriptional activation of c-fos by defining its spatial position within cells.

One of skill in the art will recognize that while the following method is applied to detection of c-fos activation, it can be applied to the analysis of any transcription factor that translocates from the cytoplasm to the nucleus upon activation. Examples of such transcription factors include, but are not limited to fos and jun homologs, NF-KB (nuclear factor kappa from B cells), NFAT (nuclear factor of activated T-lymphocytes), and STATs (signal transducer and activator of transcription) factors (For example, see Strehlow, I., and Schindler, C. 1998. *J. Biol. Chem.* 273:28049-28056; Chow, et al. 1997 *Science.* 278:1638-1641; Ding et al. 1998 *J. Biol. Chem.* 273:28897-28905; Baldwin, 1996. *Annu Rev Immunol.* 14:649-83; Kuo, C.T., and J.M. Leiden. 1999. *Annu Rev Immunol.* 17:149-87; Rao, et al. 1997. *Annu Rev Immunol.* 15:707-47; Masuda,et al. 1998. *Cell Signal.* 10:599-611; Hoey, T., and U. Schindler. 1998. *Curr Opin Genet Dev.* 8:582-7; Liu, et al. 1998. *Curr Opin Immunol.* 10:271-8.)

Thus, in this aspect of the invention, indicator cells are treated with test compounds and the distribution of luminescently labeled transcription factor is measured in space and time using a cell screening system, such as the one disclosed above. The luminescently labeled transcription factor may be expressed by or added to the cells either before, together with, or after contacting the cells with a test compound.

For example, the transcription factor may be expressed as a luminescently labeled protein chimera by transfected indicator cells. Alternatively, the luminescently labeled transcription factor may be expressed, isolated, and bulk-loaded into the indicator cells as described above, or the transcription factor may be luminescently labeled after isolation. As a further alternative, the transcription factor is expressed by

In a further aspect, kits are provided for analyzing transcription factor activation, comprising an antibody that specifically recognizes a transcription factor of interest,

and instructions for using the antibody for carrying out the methods described above. In a preferred embodiment, the transcription factor-specific antibody, or a secondary antibody that detects the transcription factor antibody, is luminescently labeled. In further preferred embodiments, the kit contains cells that express the transcription factor of interest, and/or the kit contains a compound that is known to modify activation of the transcription factor of interest, including but not limited to platelet derived growth factor (PDGF) and serum, which both modify fos activation; and interleukin 1(IL-1) and tumor necrosis factor (TNF), which both modify NF-KB activation.

In another embodiment, the kit comprises a recombinant expression vector comprising a nucleic acid encoding a transcription factor of interest that translocates from the cytoplasm to the nucleus upon activation, and instructions for using the expression vector to identify compounds that modify transcription factor activation in a cell of interest. Alternatively, the kits contain a purified, luminescently labeled transcription factor. In a preferred embodiment, the transcription factor is expressed as a fusion protein with a luminescent protein, including but not limited to green fluorescent protein, luceriferase, or mutants or fragments thereof. In various preferred embodiments, the kit further contains cells that are transfected with the expression vector, an antibody or fragment that specifically bind to the transcription factor of interest, and/or a compound that is known to modify activation of the transcription factor of interest (as above).

b. Protein Kinases

The cytoplasm to nucleus screening methods can also be used to analyze the activation of any protein kinase that is present in an inactive state in the cytoplasm and is transported to the nucleus upon activation, or that phosphorylates a substrate that translocates from the cytoplasm to the nucleus upon phosphorylation. Examples of appropriate protein kinases include, but are not limited to extracellular signal-regulated protein kinases (ERKs), c-Jun amino-terminal kinases (JNKs), Fos regulating protein

... mitogen activated protein kinase kinases (MAPKKs) ... For example, see Hall, et al. 1999. *J Biol Chem.* 274:376-83; Han, et al. 1995. *Biochim. Biophys. Acta.* 1265:224-227; Jaaro et al. 1997. *Proc. Natl. Acad. Sci. U.S.A.* 94:3742-3747; Taylor, et

al. 1994. *J. Biol. Chem.* 269:308-318; Zhao, Q., and F. S. Lee. 1999. *J Biol Chem.* 274:8355-8; Paolilloet al. 1999. *J Biol Chem.* 274:6546-52; Coso et al. 1995. Cell 81:1137-1146; Tibbles, L.A., and J.R. Woodgett. 1999. *Cell Mol Life Sci.* 55:1230-54; Schaeffer, H.J., and M.J. Weber. 1999. *Mol Cell Biol.* 19:2435-44.)

Alternatively, protein kinase activity is assayed by monitoring translocation of a luminescently labeled protein kinase substrate from the cytoplasm to the nucleus after being phosphorylated by the protein kinase of interest. In this embodiment, the substrate is non-phosphorylated and cytoplasmic prior to phosphorylation, and is translocated to the nucleus upon phosphorylation by the protein kinase. There is no requirement that the protein kinase itself translocates from the cytoplasm to the nucleus in this embodiment. Examples of such substrates (and the corresponding protein kinase) include, but are not limited to c-jun (JNK substrate); fos (FRK substrate), and p38 (p38 MAPK substrate).

Thus, in these embodiments, indicator cells are treated with test compounds and the distribution of luminescently labeled protein kinase or protein kinase substrate is measured in space and time using a cell screening system, such as the one disclosed above. The luminescently labeled protein kinase or protein kinase substrate may be expressed by or added to the cells either before, together with, or after contacting the cells with a test compound. For example, the protein kinase or protein kinase substrate may be expressed as a luminescently labeled protein chimera by transfected indicator cells. Alternatively, the luminescently labeled protein kinase or protein kinase substrate may be expressed, isolated, and bulk-loaded into the indicator cells as described above, or the protein kinase or protein kinase substrate may be luminescently labeled after isolation. As a further alternative, the protein kinase or protein kinase substrate is expressed by the indicator cell, which is subsequently contacted with a luminescent label, such as a labeled antibody, that detects the protein kinase or protein kinase substrate.

In a further embodiment, protein kinase activity is assayed by monitoring the

[illegible] requirement that either the protein kinase or the protein kinase substrate translocate from the cytoplasm to the nucleus upon activation. In a preferred embodiment, phosphorylation state is monitored by

contacting the cells with an antibody that binds only to the phosphorylated form of the protein kinase substrate of interest (For example, as disclosed in U.S. Patent No. 5,599,681).

In another preferred embodiment, a biosensor of phosphorylation is used. For example, a luminescently labeled protein or fragment thereof can be fused to a protein that has been engineered to contain (a) a phosphorylation site that is recognized by a protein kinase of interest; and (b) a nuclear localization signal that is unmasked by the phosphorylation. Such a biosensor will thus be translocated to the nucleus upon phosphorylation, and its translocation can be used as a measure of protein kinase activation.

In another aspect, kits are provided for analyzing protein kinase activation, comprising a primary antibody that specifically binds to a protein kinase, a protein kinase substrate, or a phosphorylated form of the protein kinase substrate of interest and instructions for using the primary antibody to identify compounds that modify protein kinase activation in a cell of interest. In a preferred embodiment, the primary antibody, or a secondary antibody that detects the primary antibody, is luminescently labeled. In other preferred embodiments, the kit further comprises cells that express the protein kinase of interest, and/or a compound that is known to modify activation of the protein kinase of interest, including but not limited to dibutyryl cAMP (modifies PKA), forskolin (PKA), and anisomycin (p38MAPK).

Alternatively, the kits comprise an expression vector encoding a protein kinase or a protein kinase substrate of interest that translocates from the cytoplasm to the nucleus upon activation and instructions for using the expression vector to identify compounds that modify protein kinase activation in a cell of interest. Alternatively, the kits contain a purified, luminescently labeled protein kinase or protein kinase substrate. In a preferred embodiment, the protein kinase or protein kinase substrate of interest is expressed as a fusion protein with a luminescent protein. In further preferred embodiments, the kit further comprises cells that are transfected with the expressed protein kinase substrate of interest, and/or a compound that is known to modify activation of the protein kinase of interest. (as above)

In another aspect, the present invention comprises a machine readable storage medium comprising a program containing a set of instructions for causing a cell screening system to execute the methods disclosed for analyzing transcription factor or protein kinase activation, wherein the cell screening system comprises an optical system with a stage adapted for holding a plate containing cells, a digital camera, a means for directing fluorescence or luminescence emitted from the cells to the digital camera, and a computer means for receiving and processing the digital data from the digital camera.

Example 2 Automated Screen for Compounds that Modify Cellular Morphology

Changes in cell size are associated with a number of cellular conditions, such as hypertrophy, cell attachment and spreading, differentiation, growth and division, necrotic and programmed cell death, cell motility, morphogenesis, tube formation, and colony formation.

For example, cellular hypertrophy has been associated with a cascade of alterations in gene expression and can be characterized in cell culture by an alteration in cell size, that is clearly visible in adherent cells growing on a coverslip.

Cell size can also be measured to determine the attachment and spreading of adherent cells. Cell spreading is the result of selective binding of cell surface receptors to substrate ligands and subsequent activation of signaling pathways to the cytoskeleton. Cell attachment and spreading to substrate molecules is an important step for the metastasis of cancer cells, leukocyte activation during the inflammatory response, keratinocyte movement during wound healing, and endothelial cell movement during angiogenesis. Compounds that affect these surface receptors, signaling pathways, or the cytoskeleton will affect cell spreading and can be screened by measuring cell size.

Total cellular area can be monitored by labeling the entire cell body or the cell cytoplasm using cytoskeletal markers, cytosolic volume markers, or cell surface

Missouri)) include the following:

CELL SIZE AND AREA MARKERS
Cytoskeletal Markers
• ALEXA™ 488 phalloidin (Molecular Probes, Oregon)
• Tubulin-green fluorescent protein chimeras
• Cytokeratin-green fluorescent protein chimeras
• Antibodies to cytoskeletal proteins
Cytosolic Volume Markers
• Green fluorescent proteins
• Chloromethylfluorescein diacetate (CMFDA)
• Calcein green
• BCECF/AM ester
• Rhodamine dextran
Cell Surface Markers for Lipid, Protein, or Oligosaccharide
• Dihexadecyl tetramethylindocarbocyanine perchlorate (DiIC16) lipid dyes
• Triethylammonium propyl dibutylamino styryl pyridinium (FM 4-64, FM 1-43) lipid dyes
• MITOTRACKER™ Green FM
• Lectins to oligosaccarides such as fluorescein concanavalin A or wheat germ agglutinin
• SYPRO™ Red non-specific protein markers
• Antibodies to various surface proteins such as epidermal growth factor
• Biotin labeling of surface proteins followed by fluorescent strepavidin labeleing

Protocols for cell staining with these various agents are well known to those skilled in the art. Cells are stained live or after fixation and the cell area can be measured. For example, live cells stained with DiIC16 have homogeneously labeled plasma membranes, and the projected cross-sectional area of the cell is uniformly discriminated from background by fluorescence intensity of the dye. Live cells stained with cytosolic stains such as CMFDA produce a fluorescence intensity that is proportional to cell thickness. Although cell labeling is dimmer in thin regions of the cell, total cell area can be discriminated from background. Fixed cells can be stained with cytoskeletal markers such as ALEXA™ 488 phalloidin that label polymerized actin. Phalloidin does not homogeneously stain the cytoplasm, but still permits discrimination of the total cell area from background.

Cellular hypertrophy

A screen to analyze cellular hypertrophy is implemented using the following strategy. Primary rat myocytes can be cultured in 96 well plates, treated with various compounds and then fixed and labeled with a fluorescent marker for the cell membrane or cytoplasm, or cytoskeleton, such as an antibody to a cell surface marker or a fluorescent marker for the cytoskeleton like rhodamine-phalloidin, in combination with a DNA label like Hoechst.

After focusing on the Hoechst labeled nuclei, two images are acquired, one of the Hoechst labeled nuclei and one of the fluorescent cytoplasm image. The nuclei are identified by thresholding to create a mask and then comparing the morphological descriptors of the mask with a set of user defined descriptor values. Each non-nucleus image (or "cytoplasmic image") is then processed separately. The original cytoplasm image can be thresholded, creating a cytoplasmic mask image. Local regions containing cells are defined around the nuclei. The limits of the cells in those regions are then defined by a local dynamic threshold operation on the same region in the fluorescent antibody image. A sequence of erosions and dilations is used to separate slightly touching cells and a second set of morphological descriptors is used to identify single cells. The area of the individual cells is tabulated in order to define the distribution of cell sizes for comparison with size data from normal and hypertrophic cells.

Responses from entire 96-well plates (measured as average cytoplasmic area/cell) were analyzed by the above methods, and the results demonstrated that the assay will perform the same on a well-to-well, plate-to-plate, and day-to-day basis (below a 15% cov for maximum signal). The data showed very good correlation for each day, and that there was no variability due to well position in the plate.

The following totals can be computed for the field. The aggregate whole nucleus area is the number of nonzero pixels in the nuclear mask. The average whole nucleus area is the aggregate whole nucleus area divided by the total number of nuclei. For each cytoplasm image several values can be computed. These are the total

[illegible] intensities [illegible] pixels [illegible] cytoplasmic mask. The cytoplasmic area per nucleus is the total cytoplasmic area divided by the total nucleus count. The cytoplasmic intensity per nucleus is the

aggregate cytoplasm intensity divided by the total nucleus count. The average cytoplasm intensity is the aggregate cytoplasm intensity divided by the cytoplasm area. The cytoplasm nucleus ratio is the total cytoplasm area divided by the total nucleus area.

Additionally, one or more fluorescent antibodies to other cellular proteins, such as the major muscle proteins actin or myosin, can be included. Images of these additional labeled proteins can be acquired and stored with the above images, for later review, to identify anomalies in the distribution and morphology of these proteins in hypertrophic cells. This example of a multi-parametric screen allows for simultaneous analysis of cellular hypertrophy and changes in actin or myosin distribution.

One of skill in the art will recognize that while the example analyzes myocyte hypertrophy, the methods can be applied to analyzing hypertrophy, or general morphological changes in any cell type.

Cell morphology assays for prostate carcinoma

Cell spreading is a measure of the response of cell surface receptors to substrate attachment ligands. Spreading is proportional to the ligand concentration or to the concentration of compounds that reduce receptor-ligand function. One example of selective cell-substrate attachment is prostate carcinoma cell adhesion to the extracellular matrix protein collagen. Prostate carcinoma cells metastasize to bone via selective adhesion to collagen.

Compounds that interfere with metastasis of prostate carcinoma cells were screened as follows. PC3 human prostate carcinoma cells were cultured in media with appropriate stimulants and are passaged to collagen coated 96 well plates. Ligand concentration can be varied or inhibitors of cell spreading can be added to the wells. Examples of compounds that can affect spreading are receptor antagonists such as integrin- or proteoglycan-blocking antibodies, signaling inhibitors including phosphatidyl inositol-3 kinase inhibitors, and cytoskeletal inhibitors such as

[illegible] hypertrophy. The size of cells under these various conditions, as measured by cytoplasmic staining, can be distinguished above background levels. The number of

cells per field is determined by measuring the number of nuclei stained with the Hoechst DNA dye. The area per cell is found by dividing the cytoplasmic area (phalloidin image) by the cell number (Hoechst image). The size of cells is proportional to the ligand-receptor function. Since the area is determined by ligand concentration and by the resultant function of the cell, drug efficacy, as well as drug potency, can be determined by this cell-based assay. Other measurements can be made as discussed above for cellular hypertrophy.

The methods for analyzing cellular morphology can be used in a combined high throughput-high content screen. In one example, the high throughput mode scans the whole well for an increase in fluorescent phalloidin intensity. A threshold is set above which both nuclei (Hoechst) and cells (phalloidin) are measured in a high content mode. In another example, an environmental biosensor (examples include, but are not limited to, those biosensors that are sensitive to calcium and pH changes) is added to the cells, and the cells are contacted with a compound. The cells are scanned in a high throughput mode, and those wells that exceed a pre-determined threshold for luminescence of the biosensor are scanned in a high content mode.

In a further aspect, kits are provided for analyzing cellular morphology, comprising a luminescent compound that can be used to specifically label the cell cytoplasm, membrane, or cytoskeleton (such as those described above), and instructions for using the luminescent compound to identify test stimuli that induce or inhibit changes in cellular morphology according to the above methods. In a preferred embodiment, the kit further comprises a luminescent marker for cell nuclei. In a further preferred embodiment, the kit comprises at least one compound that is known to modify cellular morphology, including, but not limited to integrin- or proteoglycan-blocking antibodies, signaling inhibitors including phosphatidyl inositol-3 kinase inhibitors, and cytoskeletal inhibitors such as cytochalasin D.

In another aspect, the present invention comprises a machine readable storage medium comprising a program containing a set of instructions for causing a cell

[illegible] system with a stage adapted to holding a plate containing cells, a digital camera, a means for directing fluorescence or

luminescence emitted from the cells to the digital camera, and a computer means for receiving and processing the digital data from the digital camera.

Example 3 Dual Mode High Throughput and High-Content Screen

The following example is a screen for activation of a G-protein coupled receptor (GPCR) as detected by the translocation of the GPCR from the plasma membrane to a proximal nuclear location. This example illustrates how a high throughput screen can be coupled with a high-content screen in the dual mode System for Cell Based Screening.

G-protein coupled receptors are a large class of 7 trans-membrane domain cell surface receptors. Ligands for these receptors stimulate a cascade of secondary signals in the cell, which may include, but are not limited to, Ca^{++} transients, cyclic AMP production, inositol triphosphate (IP_3) production and phosphorylation. Each of these signals are rapid, occuring in a matter of seconds to minutes, but are also generic. For example, many different GPCRs produce a secondary Ca^{++} signal when activated. Stimulation of a GPCR also results in the transport of that GPCR from the cell surface membrane to an internal, proximal nuclear compartment. This internalization is a much more receptor-specific indicator of activation of a particular receptor than are the secondary signals described above.

Figure 19 illustrates a dual mode screen for activation of a GPCR. Cells carrying a stable chimera of the GPCR with a blue fluorescent protein (BFP) would be loaded with the acetoxymethylester form of Fluo-3, a cell permeable calcium indicator (green fluorescence) that is trapped in living cells by the hydrolysis of the esters. They would then be deposited into the wells of a microtiter plate 601. The wells would then be treated with an array of test compounds using a fluid delivery system, and a short sequence of Fluo-3 images of the whole microtiter plate would be acquired and analyzed for wells exhibiting a calcium response (i.e., high throughput mode). The

[illegible] brightly due to the [illegible] released upon stimulation of the receptors. The locations of wells containing compounds that induced a response 602, would then be transferred

to the HCS program and the optics switched for detailed cell by cell analysis of the blue fluorescence for evidence of GPCR translocation to the perinuclear region. The bottom of Figure 19 illustrates the two possible outcomes of the analysis of the high resolution cell data. The camera images a sub-region 604 of the well area 603, producing images of the fluorescent cells 605. In well C4, the uniform distribution of the fluorescence in the cells indicates that the receptor has not internalized, implying that the Ca^{++} response seen was the result of the stimulation of some other signalling system in the cell. The cells in well E9 606 on the other hand, clearly indicate a concentration of the receptor in the perinuclear region clearly indicating the full activation of the receptor. Because only a few hit wells have to be analyzed with high resolution, the overall throughput of the dual mode system can be quite high, comparable to the high throughput system alone.

Example 4 Kinetic High Content Screen

The following is an example of a screen to measure the kinetics of internalization of a receptor. As described above, the stimulation of a GPCR, results in the internalization of the receptor, with a time course of about 15 min. Simply detecting the endpoint as internalized or not, may not be sufficient for defining the potency of a compound as a GPCR agonist or antagonist. However, 3 time points at 5 min intervals would provide information not only about potency during the time course of measurement, but would also allow extrapolation of the data to much longer time periods. To perform this assay, the sub-region would be defined as two rows, the sampling interval as 5 minutes and the total number of time points 3. The system would then start by scanning two rows, and then adding reagent to the two rows, establishing the time=0 reference. After reagent addition, the system would again scan the two row sub-region acquiring the first time point data. Since this process would take about 250 seconds, including scanning back to the beginning of the sub-region, the second 2 row sub-region. The final two 2-row sub-regions would be scanned to finish all the wells on the plate, resulting in four time points for each well over the whole

plate. Although the time points for the wells would be offset slightly relative to time=0, the spacing of the time points would be very close to the required 5 minutes, and the actual acquisition times and results recorded with much greater precision than in a fixed-cell screen.

Example 5 High-content screen of human glucocorticoid receptor translocation

One class of HCS involves the drug-induced dynamic redistribution of intracellular constituents. The human glucocorticoid receptor (hGR), a single "sensor" in the complex environmental response machinery of the cell, binds steroid molecules that have diffused into the cell. The ligand-receptor complex translocates to the nucleus where transcriptional activation occurs (Htun et al., *Proc. Natl. Acad. Sci.* 93:4845, 1996).

In general, hormone receptors are excellent drug targets because their activity lies at the apex of key intracellular signaling pathways. Therefore, a high-content screen of hGR translocation has distinct advantage over *in vitro* ligand-receptor binding assays. The availability of up to two more channels of fluorescence in the cell screening system of the present invention permits the screen to contain two additional parameters in parallel, such as other receptors, other distinct targets or other cellular processes.

Plasmid construct. A eukaryotic expression plasmid containing a coding sequence for a green fluorescent protein – human glucocorticoid receptor (GFP-hGR) chimera was prepared using GFP mutants (Palm et al., *Nat. Struct. Biol.* 4:361 (1997). The construct was used to transfect a human cervical carcinoma cell line (HeLa).

Cell preparation and transfection. HeLa cells (ATCC CCL-2) were trypsinized and plated using DMEM containing 5% charcoal/dextran-treated fetal bovine serum (FBS) (HyClone) and 1% penicillin-streptomycin (C-DMEM) 12-24 hours prior to transfection and incubated at 37°C and 5% CO_2 . Transfections were performed by calcium phosphate co-precipitation (Graham and Van der Eb, *Virology* 52:456, 1973

[illegible] Spring Harbor [illegible] with Lipofectamine [illegible] Technologies, Gaithersburg, MD). For the calcium phosphate transfections, the medium was replaced, prior to transfection, with DMEM containing 5%

charcoal/dextran-treated FBS. Cells were incubated with the calcium phosphate-DNA precipitate for 4-5 hours at 37°C and 5% CO_2, washed 3-4 times with DMEM to remove the precipitate, followed by the addition of C-DMEM.

Lipofectamine transfections were performed in serum-free DMEM without antibiotics according to the manufacturer's instructions (Life Technologies, Gaithersburg, MD). Following a 2-3 hour incubation with the DNA-liposome complexes, the medium was removed and replaced with C-DMEM. All transfected cells in 96-well microtiter plates were incubated at 33°C and 5% CO_2 for 24-48 hours prior to drug treatment. Experiments were performed with the receptor expressed transiently in HeLa cells.

Dexamethasone induction of GFP-hGR translocation. To obtain receptor-ligand translocation kinetic data, nuclei of transfected cells were first labeled with 5 µg/ml Hoechst 33342 (Molecular Probes) in C-DMEM for 20 minutes at 33°C and 5% CO_2. Cells were washed once in Hank's Balanced Salt Solution (HBSS) followed by the addition of 100 nM dexamethasone in HBSS with 1% charcoal/dextran-treated FBS. To obtain fixed time point dexamethasone titration data, transfected HeLa cells were first washed with DMEM and then incubated at 33°C and 5% CO_2 for 1 h in the presence of 0 – 1000 nM dexamethasone in DMEM containing 1% charcoal/dextran-treated FBS. Cells were analyzed live or they were rinsed with HBSS, fixed for 15 min with 3.7% formaldehyde in HBSS, stained with Hoechst 33342, and washed before analysis. The intracellular GFP-hGR fluorescence signal was not diminished by this fixation procedure.

Image acquisition and analysis. Kinetic data were collected by acquiring fluorescence image pairs (GFP-hGR and Hoechst 33342-labeled nuclei) from fields of living cells at 1 min intervals for 30 min after the addition of dexamethasone. Likewise, image pairs were obtained from each well of the fixed time point screening plates 1 h after the addition of dexamethasone. In both cases, the image pairs obtained at each time point were used to define nuclear and cytoplasmic regions in each cell

integrated fluorescence intensity of the chimera in the cytoplasm or as a nuclear-cytoplasmic difference of GFP fluorescence. In the fixed time point screen this translocation ratio was calculated from data obtained

from at least 200 cells at each concentration of dexamethasone tested. Drug-induced translocation of GFP-hGR from the cytoplasm to the nucleus was therefore correlated with an increase in the translocation ratio.

Results. Figure 20 schematically displays the drug-induced cytoplasm 253 to nucleus 252 translocation of the human glucocorticoid receptor. The upper pair of schematic diagrams depicts the localization of GFP-hGR within the cell before 250 (A) and after 251 (B) stimulation with dexamethasone. Under these experimental conditions, the drug induces a large portion of the cytoplasmic GFP-hGR to translocate into the nucleus. This redistribution is quantified by determining the integrated intensities ratio of the cytoplasmic and nuclear fluorescence in treated 255 and untreated 254 cells. The lower pair of fluorescence micrographs show the dynamic redistribution of GFP-hGR in a single cell, before 254 and after 255 treatment. The HCS is performed on wells containing hundreds to thousands of transfected cells and the translocation is quantified for each cell in the field exhibiting GFP fluorescence. Although the use of a stably transfected cell line would yield the most consistently labeled cells, the heterogeneous levels of GFP-hGR expression induced by transient transfection did not interfere with analysis by the cell screening system of the present invention.

To execute the screen, the cell screening system scans each well of the plate, images a population of cells in each, and analyzes cells individually. Here, two channels of fluorescence are used to define the cytoplasmic and nuclear distribution of the GFP-hGR within each cell. Depicted in Figure 21 is the graphical user interface of the cell screening system near the end of a GFP-hGR screen. The user interface depicts the parallel data collection and analysis capability of the system. The windows labeled "Nucleus" 261 and "GFP-hGR" 262 show the pair of fluorescence images being obtained and analyzed in a single field. The window labeled "Color Overlay" 260 is formed by pseudocoloring the above images and merging them so the user can immediately identify cellular changes. Within the "Stored Object Regions" window

[illegible] Furthermore, as the HCS data are being collected, they are analyzed, in this case for GFP-hGR translocation, and translated into an immediate "hit" response. The 96 well plate depicted in the lower window of the screen 267 shows which wells have

met a set of user-defined screening criteria. For example, a white-colored well 269 indicates that the drug-induced translocation has exceeded a predetermined threshold value of 50%. On the other hand, a black-colored well 270 indicates that the drug being tested induced less than 10% translocation. Gray-colored wells 268 indicate "hits" where the translocation value fell between 10% and 50%. Row "E" on the 96 well plate being analyzed 266 shows a titration with a drug known to activate GFP-hGR translocation, dexamethasone. This example screen used only two fluorescence channels. Two additional channels (Channels 3 263 and 4 264) are available for parallel analysis of other specific targets, cell processes, or cytotoxicity to create multiple parameter screens.

There is a link between the image database and the information database that is a powerful tool during the validation process of new screens. At the completion of a screen, the user has total access to image and calculated data (Figure 22). The comprehensive data analysis package of the cell screening system allows the user to examine HCS data at multiple levels. Images 276 and detailed data in a spread sheet 279 for individual cells can be viewed separately, or summary data can be plotted. For example, the calculated results of a single parameter for each cell in a 96 well plate are shown in the panel labeled Graph 1 275. By selecting a single point in the graph, the user can display the entire data set for a particular cell that is recalled from an existing database. Shown here are the image pair 276 and detailed fluorescence and morphometric data from a single cell (Cell #118, gray line 277). The large graphical insert 278 shows the results of dexamethasone concentration on the translocation of GFP-hGR. Each point is the average of data from at least 200 cells. The calculated EC_{50} for dexamethasone in this assay is 2 nM.

A powerful aspect of HCS with the cell screening system is the capability of kinetic measurements using multicolor fluorescence and morphometric parameters in living cells. Temporal and spatial measurements can be made on single cells within a population of cells in a field. Figure 23 shows kinetic data for the dexamethasone

[illegible] were treated with 1 nM dexamethasone and the translocation of GFP-hGR was measured over time in a population of single cells. The graph shows the response of transfected cells 285, 286, 287, and 288 and non-

transfected cells 289. These data also illustrate the ability to analyze cells with different expression levels.

Example 6 High-content screen of drug-induced apoptosis

Apoptosis is a complex cellular program that involves myriad molecular events and pathways. To understand the mechanisms of drug action on this process, it is essential to measure as many of these events within cells as possible with temporal and spatial resolution. Therefore, an apoptosis screen that requires little cell sample preparation yet provides an automated readout of several apoptosis-related parameters would be ideal. A cell-based assay designed for the cell screening system has been used to simultaneously quantify several of the morphological, organellar, and macromolecular hallmarks of paclitaxel-induced apoptosis.

Cell preparation. The cells chosen for this study were mouse connective tissue fibroblasts (L-929; ATCC CCL-1) and a highly invasive glioblastoma cell line (SNB-19; ATCC CRL-2219) (Welch et al., *In Vitro Cell. Dev. Biol.* 31:610, 1995). The day before treatment with an apoptosis inducing drug, 3500 cells were placed into each well of a 96-well plate and incubated overnight at 37°C in a humidified 5% CO_2 atmosphere. The following day, the culture medium was removed from each well and replaced with fresh medium containing various concentrations of paclitaxel (0 – 50 μM) from a 20 mM stock made in DMSO. The maximal concentration of DMSO used in these experiments was 0.25%. The cells were then incubated for 26 h as above. At the end of the paclitaxel treatment period, each well received fresh medium containing 750 nM MitoTracker Red (Molecular Probes; Eugene, OR) and 3 μg/ml Hoechst 33342 DNA-binding dye (Molecular Probes) and was incubated as above for 20 min. Each well on the plate was then washed with HBSS and fixed with 3.7% formaldehyde in HBSS for 15 min at room temperature. The formaldehyde was washed out with HBSS and the cells were permeabilized for 90 s with 0.5% (v/v) Triton X-100, washed with

[illegible] and the plate stored at [illegible] necessary. The fluorescence signals from plates stored this way were stable for at least two weeks after preparation. As in the nuclear

translocation assay, fluorescence reagents can be designed to convert this assay into a live cell high-content screen.

Image acquisition and analysis on the ArrayScan System. The fluorescence intensity of intracellular MitoTracker Red, Hoechst 33342, and Bodipy FL phallacidin was measured with the cell screening system as described *supra*. Morphometric data from each pair of images obtained from each well was also obtained to detect each object in the image field (*e.g.*, cells and nuclei), and to calculate its size, shape, and integrated intensity.

Calculations and output. A total of 50-250 cells were measured per image field. For each field of cells, the following calculations were performed: (1) The average nuclear area (μm^2) was calculated by dividing the total nuclear area in a field by the number of nuclei detected. (2) The average nuclear perimeter (μm) was calculated by dividing the sum of the perimeters of all nuclei in a field by the number of nuclei detected in that field. Highly convoluted apoptotic nuclei had the largest nuclear perimeter values. (3) The average nuclear brightness was calculated by dividing the integrated intensity of the entire field of nuclei by the number of nuclei in that field. An increase in nuclear brightness was correlated with increased DNA content. (4) The average cellular brightness was calculated by dividing the integrated intensity of an entire field of cells stained with MitoTracker dye by the number of nuclei in that field. Because the amount of MitoTracker dye that accumulates within the mitochondria is proportional to the mitochondrial potential, an increase in the average cell brightness is consistent with an increase in mitochondrial potential. (5) The average cellular brightness was also calculated by dividing the integrated intensity of an entire field of cells stained with Bodipy FL phallacidin dye by the number of nuclei in that field. Because the phallotoxins bind with high affinity to the polymerized form of actin, the amount of Bodipy FL phallacidin dye that accumulates within the cell is proportional to actin polymerization state. An increase in the average cell brightness is consistent with an increase in actin polymerization

[illegible] cells [illegible] increasing amounts of paclitaxel caused nuclei to enlarge and fragment 293, a hallmark of apoptosis. Quantitative analysis of these and other images obtained by the cell screening system is presented in the same figure.

Each parameter measured showed that the L-929 cells 296 were less sensitive to low concentrations of paclitaxel than were SNB-19 cells 297. At higher concentrations though, the L-929 cells showed a response for each parameter measured. The multiparameter approach of this assay is useful in dissecting the mechanisms of drug action. For example, the area, brightness, and fragmentation of the nucleus 298 and actin polymerization values 294 reached a maximum value when SNB-19 cells were treated with 10 nM paclitaxel (Figure 24; top and bottom graphs). However, mitochondrial potential 295 was minimal at the same concentration of paclitaxel (Figure 24; middle graph). The fact that all the parameters measured approached control levels at increasing paclitaxel concentrations (>10 nM) suggests that SNB-19 cells have low affinity drug metabolic or clearance pathways that are compensatory at sufficiently high levels of the drug. Contrasting the drug sensitivity of SNB-19 cells 297, L-929 showed a different response to paclitaxel 296. These fibroblastic cells showed a maximal response in many parameters at 5 μM paclitaxel, a 500-fold higher dose than SNB-19 cells. Furthermore, the L-929 cells did not show a sharp decrease in mitochondrial potential 295 at any of the paclitaxel concentrations tested. This result is consistent with the presence of unique apoptosis pathways between a normal and cancer cell line. Therefore, these results indicate that a relatively simple fluorescence labeling protocol can be coupled with the cell screening system of the present invention to produce a high-content screen of key events involved in programmed cell death.

Background

A key to the mechanism of apoptosis was the discovery that, irrespective of the lethal stimulus, death results in identical apoptotic morphology that includes cell and organelle dismantling and repackaging, DNA cleavage to nucleosome sized fragments, and engulfment of the fragmented cell to avoid an inflammatory response. Apoptosis is therefore distinct from necrosis, which is mediated more by acute trauma to a cell,

[illegible] criteria for determining whether a cell is undergoing apoptosis (Wyllie et al. 1980. *Int Rev Cytol.* 68:251-306; Thompson, 1995. *Science.* 267:1456-62; Majno

and Joris. 1995. *Am J Pathol.* 146:3-15; Allen et al. 1998. *Cell Mol Life Sci.* 54:427-45) include distinct morphological changes in the appearance of the cell, as well as alterations in biochemical and molecular markers. For example, apoptotic cells often undergo cytoplasmic membrane blebbing, their chromosomes rapidly condense and aggregate around the nuclear periphery, the nucleus fragments, and small apoptotic bodies are formed. In many, but not all, apoptotic cells, chromatin becomes a target for specific nucleases that cleave the DNA.

Apoptosis is commonly accompanied by a characteristic change in nuclear morphology (chromatin condensation or fragmentation) and a step-wise fragmentation of DNA culminating in the formation of mono- and/or oligomeric fragments of 200 base pairs. Specific changes in organellar function, such as mitochondrial membrane potential, occur. In addition, specific cysteine proteases (caspases) are activated, which catalyzes a highly selective pattern of protein degradation by proteolytic cleavage after specific aspartic acid residues. In addition, the external surface exposure of phosphatidylserine residues (normally on the inner membrane leaflet) allows for the recognition and elimination of apoptotic cells, before the membrane breaks up and cytosol or organelles spill into the intercellular space and elicit inflammatory reactions. Moreover, cells undergoing apoptosis tend to shrink, while also having a reduced intracellular potassium level.

The general patterns of apoptotic signals are very similar among different cell types and apoptotic inducers. However, the details of the pathways actually vary significantly depending on cell type and inducer. The dependence and independence of various signal transduction pathways involved in apoptosis are currently topics of intense research. We show here that the pathway also varies depending upon the dose of the inducer in specific cell types.

Nuclear Morphology

Cells undergoing apoptosis usually [illegible]

[illegible] response in a given cell type appears [illegible] depending on the apoptotic inducer. During nuclear fragmentation, a circular or oval nucleus becomes increasingly lobular. Eventually, the nucleus fragments dramatically into multiple sub-nuclei. Sometimes the

density of the chromatin within the lobular nucleus may show spatial variations in distribution (heterochromatization), approximating the margination seen in nuclear condensation.

Nuclear condensation has been reported in some cell types, such as MCF-7 (Saunders et al. 1997. *Int J Cancer*. 70:214-20). Condensation appears to arise as a consequence of the loss of structural integrity of the euchromatin, nuclear matrix and nuclear lamina (Hendzel et al. 1998. *J Biol Chem*. 273:24470-8). During nuclear condensation, the chromatin concentrates near the margin of the nucleus, leading to the overall shrinkage of the nucleus. Thus, the use of nuclear morphology as a measure of apoptosis must take both condensation and fragmentation into account.

Material and Methods

Cells were plated into 96-well plates at densities of 3×10^3 to 1×10^4 cells/well. The following day apoptotic inducers were added at indicated concentrations and cells were incubated for indicated time periods (usually 16-30 hours). The next day medium was removed and cells were stained with 5 µg/ml Hoechst (Molecular Probes, Inc.) in fresh medium and incubated for 30 minutes at 37°C. Cells were washed in Hank's Balanced Salt Solution (HBSS) and fixed with 3.7% formaldehyde in HBSS at room temperature. Cells were washed 2X with HBSS at room temperature and the plate was sealed.

Quantitation of changes in nuclear morphology upon induction of apoptosis was accomplished by (1) measuring the effective size of the nuclear region; and (2) measuring the degree of convolution of the perimeter. The size parameter provides the more sensitive measure of nuclear condensation, whereas the perimeter measure provides a more sensitive measure of nuclear fragmentation.

Results & Discussion

[illegible] illustrated a slightly more complicated, yet clearly visible response. Staurosporine appeared to stimulate nuclear condensation at lower doses and nuclear fragmentation at higher doses (**Fig 25C and 25D**). In contrast, paclitaxel induced a consistent increase in

nuclear fragmentation with increasing concentrations. The response of MCF-7 cells varied dramatically depending upon the apoptotic inducer. Staurosporine appeared to elicit nuclear condensation whereas paclitaxel induced nuclear fragmentation (**Fig 25E and 25F**).

Figure 26 illustrates the dose response of cells in terms of both nuclear size and nuclear perimeter convolution. There appears to be a swelling of the nuclei that precedes the fragmentation.

Result of evaluation: Differential responses by cell lines and by apoptotic inducers were observed in a dose dependent manner, indicating that this assay will be useful for detecting changes in the nucleus characteristic of apoptosis.

Actin reorganization

We assessed changes in the actin cytoskeleton as a potential parameter related to apoptotic changes. This was based on preliminary observations of an early increase in f-actin content detected with fluorescent phalloidin labeling, an f-actin specific stain (our unpublished data; Levee et al. 1996. *Am J Physiol.* 271:C1981-92; Maekawa et al. 1996. *Clin Exp Immunol.* 105:389-96). Changes in the actin cytoskeleton during apoptosis have not been observed in all cell types. (Endresen et al. 1995. *Cytometry.* 20:162-71, van Engeland et al. 1997. *Exp Cell Res.* 235:421-30).

Material and Methods

Cells were plated in 96-well plates at densities of 3×10^3 to 1×10^4 cells/well. The following day apoptotic inducers were added at indicated concentrations. Cells were incubated for the indicated time periods (usually 16-30 hours). The next day the medium was removed and cells were stained with 5 µg/ml Hoechst (Molecular Probes, Inc.) in fresh medium and incubated for 30 minutes at 30°C. Cells were washed in HBSS and fixed with 3.7% formaldehyde in HBSS at room temperature. Plates were washed with HBSS and permeabilized with 0.5% v/v Triton X-100 in HBSS at room temperature. Plates were washed in HBSS [illegible]

HBSS [illegible] and the plate was sealed.

Quantitation of f-actin content was accomplished by measuring the intensity of phalloidin staining around the nucleus. This was determined to be a reasonable

approximation of a full cytoplasmic average of the intensity. The mask used to approximate this cytoplasmic measure was derived from the nuclear mask defined by the Hoechst stain. Derivation was accomplished by combinations of erosions and dilations.

Results and Discussion

Changes in f-actin content varied based on cell type and apoptotic inducer (**Fig 27**). Staurosporine (30 hours) induced increases in f-actin in L929 (**Fig. 27A**) and BHK (**Fig. 27B**) cells. MCF-7 cells exhibited a concentration-dependent response. At low concentrations (**Fig. 27E**) there appeared to be a decrease in f-actin content. At higher concentrations, f-actin content increased. Paclitaxel (30 hours) treatment led to a wide variety of responses. L929 cells responded with graded increases in f-actin (**Fig. 27B**) whereas both BHK and MCF-7 responses were highly variable (**Figs. 27D & 27F,** respectively).

Result of Evaluation: Both increases and decreases in signal intensity were measured for several cell lines and found to exhibit a concentration dependent response. For certain cell line/apoptotic inducer pairs this could be a statistically significant apoptotic indicator.

Changes in Mitochondrial Mass/Potential

Introduction

Changes in mitochondria play a central role in apoptosis (Henkart and Grinstein. 1996. *J Exp Med.* 183:1293-5). Mitochondria release apoptogenic factors through the outer membrane and dissipate the electrochemical gradient of the inner membrane. This is thought to occur via formation of the mitochondria permeability transition (MPT), although it is apparently not true in all cases. An obvious manifestation of the formation of the MPT [illegible]

expression of the anti-apoptotic protein Bcl-2 prevents cell death, suggesting the formation of the MPT may be a rate-limiting event of the death process (For review see: Kroemer et al. 1998. *Annu Rev Physiol.* 60:619-42). It has also been observed that

mitochondria can proliferate during stimulation of apoptosis (Mancini et al. 1997. *J Cell Biol.* 138:449-69; Camilleri-Broet et al. 1998. *Exp Cell Res.* 239:277-92).

One approach for measuring apoptosis-induced changes in mitochondria is to measure the mitochondrial membrane potential. Of the methods available, the simplest measure is the redistribution of a cationic dye that distributes within intracellular organelles based on the membrane potential. Such an approach traditionally requires live cells for the measurements. The recent introduction of the MitoTracker dyes (Poot et al. 1997. *Cytometry.* 27:358-64; available from Molecular Probes, Inc., Oregon) provides a means of measuring mitochondrial membrane potential after fixation.

Given the observations of a possible increase in mitochondrial mass during apoptosis, the amount of dye labeling the mitochondria is related to both membrane potential and the number of mitochondria. If the number of mitochondria remains constant then the amount of dye is directly related to the membrane potential. If the number of mitochondria is not constant, then the signal will likely be dominated by the increase in mass (Reipert et al. 1995. *Exp Cell Res.* 221:281-8).

Probes are available that allow a clear separation between changes in mass and potential in HCS assays. Mitochondrial mass is measured directly by labeling with Mitotracker Green FM (Poot and Pierce, 1999, *Cytometry.* 35:311-7; available from Molecular Probes, Inc., Oregon). The labeling is independent of mitochondrial membrane potential but proportional to mitochondrial mass. This also provides a means of normalizing other mitochondrial measures in each cell with respect to mitochondrial mass.

Material and Methods

Cells were plated into 96-well plates at densities of 3×10^3 to 1×10^4 cells/well. The following day apoptotic inducers were added at the indicated concentrations and cells were incubated for the indicated time periods (usually 16-30 hours). Cells were stained with 5 µg/ml Hoechst (Molecular Probes

Cells were washed in HBSS and fixed with 3.7% formaldehyde in HBSS at room temperature. Plates were washed with HBSS and permeabilized with 0.5% v/v Triton X-100 in HBSS at room temperature. Cells were washed 2X with HBSS at room

temperature and the plate was sealed. For dual labeling of mitochondria, cells were treated with 200 nM Mitotracker Green and 200 nM Mitotracker Red for 0.5 hours before fixation.

Results & Discussion

Induction of apoptosis by staurosporine and paclitaxel led to varying mitochondrial changes depending upon the stimulus. L929 cells exhibited a clear increase in mitochondrial mass with increasing staurosporine concentrations (**Fig. 28**). BHK cells exhibited either a decrease in membrane potential at lower concentrations of staurosporine, or an increase in mass at higher concentrations of staurosporine (**Fig. 28C**). MCF-7 cells responded by a consistent decrease in mitochondrial membrane potential in response to increasing concentrations of staurosporine (**Fig 28E**). Increasing concentrations of paclitaxel caused consistent increases in mitochondrial mass (**Fig 28B, 28D**, and **28F**).

The mitochondrial membrane potential is measured by labeling mitochondria with both Mitotracker Green FM and Mitotracker Red (Molecular Probes, Inc). Mitotracker Red labeling is proportional to both mass and membrane potential. Mitotracker Green FM labeling is proportional to mass. The ratio of Mitotracker Red signal to the Mitotracker Green FM signal provides a measure of mitochondrial membrane potential (Poot and Pierce, 1999). This ratio normalizes the mitochondrial mass with respect to the Mitotracker Red signal. (**See Figure 28G**) Combining the ability to normalize to mitochondrial mass with a measure of the membrane potential allows independent assessment of both parameters.

Result of Evaluation: Both decreases in potential and increases in mass were observed depending on the cell line and inducer tested. Dose dependent correlation demonstrates that this is a promising apoptotic indicator.

[illegible]

[illegible] mitochondrial mass mitochondrial potential data shown earlier were collected as multiparameter assays, but were presented individually for clarity.

Example 7. Protease induced translocation of a signaling enzyme containing a disease-associated sequence from cytoplasm to nucleus.

Plasmid construct. A eukaryotic expression plasmid containing a coding sequence for a green fluorescent protein – caspase (Cohen (1997), *Biochemical J.* 326:1-16; Liang et al. (1997), *J. of Molec. Biol.* 274:291-302) chimera is prepared using GFP mutants. The construct is used to transfect eukaryotic cells.

Cell preparation and transfection. Cells are trypsinized and plated 24 h prior to transfection and incubated at 37°C and 5% CO_2. Transfections are performed by methods including, but not limited to calcium phosphate coprecipitation or lipofection. Cells are incubated with the calcium phosphate-DNA precipitate for 4-5 hours at 37°C and 5% CO_2, washed 3-4 times with DMEM to remove the precipitate, followed by the addition of C-DMEM. Lipofectamine transfections are performed in serum-free DMEM without antibiotics according to the manufacturer's instructions. Following a 2-3 hour incubation with the DNA-liposome complexes, the medium is removed and replaced with C-DMEM.

Apopototic induction of Caspase-GFP translocation. To obtain Caspase-GFP translocation kinetic data, nuclei of transfected cells are first labeled with 5 µg/ml Hoechst 33342 (Molecular Probes) in C-DMEM for 20 minutes at 37°C and 5% CO_2. Cells are washed once in Hank's Balanced Salt Solution (HBSS) followed by the addition of compounds that induce apoptosis. These compounds include, but are not limited to paclitaxel, staurosporine, ceramide, and tumor necrosis factor. To obtain fixed time point titration data, transfected cells are first washed with DMEM and then incubated at 37°C and 5% CO_2 for 1 h in the presence of 0 – 1000 nM compound in DMEM. Cells are analyzed live or they are rinsed with HBSS, fixed for 15 min with 3.7% formaldehyde in HBSS, stained with Hoechst 33342, and washed before analysis.

Image acquisition and analysis. Kinetic data are collected by acquiring fluorescence image pairs (Caspase-GFP and Hoechst 33342) [illegible]

[illegible] stained from each well of the fixed time point screening plates [illegible] after the addition of compound. In both cases, the image pairs obtained at each time point are used to define nuclear and cytoplasmic regions in each cell. Translocation of

Caspase-GFP is calculated by dividing the integrated fluorescence intensity of Caspase-GFP in the nucleus by the integrated fluorescence intensity of the chimera in the cytoplasm or as a nuclear-cytoplasmic difference of GFP fluorescence. In the fixed time point screen this translocation ratio is calculated from data obtained from at least 200 cells at each concentration of compound tested. Drug-induced translocation of Caspase-GFP from the cytoplasm to the nucleus is therefore correlated with an increase in the translocation ratio. Molecular interaction libraries including, but not limited to those comprising putative activators or inhibitors of apoptosis-activated enzymes are use to screen the indicator cell lines and identify a specific ligand for the DAS, and a pathway activated by compound activity.

Example 8. Identification of novel steroid receptors from DAS

Two sources of material and/or information are required to make use of this embodiment, which allows assessment of the function of an uncharacterized gene. First, disease associated sequence bank(s) containing cDNA sequences suitable for transfection into mammalian cells can be used. Because every RADE or differential expression experiment generates up to several hundred sequences, it is possible to generate an ample supply of DAS. Second, information from primary sequence database searches can be used to place DAS into broad categories, including, but not limited to, those that contain signal sequences, seven trans-membrane motifs, conserved protease active site domains, or other identifiable motifs. Based on the information acquired from these sources, method types and indicator cell lines to be transfected are selected. A large number of motifs are already well characterized and encoded in the linear sequences contained within the large number genes in existing genomic databases.

In one embodiment, the following steps are taken:

1) Information from the DAS identification experiment (including database

2) Sorting of DNA sequences or DAS by identifiable motifs (ie. signal sequences, 7- transmembrane domains, conserved protease active site domains, etc.) This initial grouping will determine fluorescent tagging strategies, host cell lines,

indicator cell lines, and banks of bioactive molecules to be screened, as described *supra*.

3) Using well established molecular biology methods, ligate DAS into an expression vector designed for this purpose. Generalized expression vectors contain promoters, enhancers, and terminators for which to deliver target sequences to the cell for transient expression. Such vectors may also contain antibody tagging sequences, direct association sequences, chromophore fusion sequences like GFP, etc. to facilitate detection when expressed by the host.

4) Transiently transfect cells with DAS containing vectors using standard transfection protocols including: calcium phosphate co-precipitation, liposome mediated, DEAE dextran mediated, polycationic mediated, viral mediated, or electroporation, and plate into microtiter plates or microwell arrays. Alternatively, transfection can be done directly in the microtiter plate itself.

5) Carry out the cell screening methods as described *supra*.

In this embodiment, DAS shown to possess a motif(s) suggestive of transcriptional activation potential (for example, DNA binding domain, amino terminal modulating domain, hinge region, or carboxy terminal ligand binding domain) are utilized to identify novel steroid receptors.

Defining the fluorescent tags for this experiment involves identification of the nucleus through staining, and tagging the DAS by creating a GFP chimera via insertion of DAS into an expression vector, proximally fused to the gene encoding GFP. Alternatively, a single chain antibody fragment with high affinity to some portion of the expressed DAS could be constructed using technology available in the art (Cambridge Antibody Technologies) and linked to a fluorophore (FITC) to tag the putative transcriptional activator/receptor in the cells. This alternative would provide an external tag requiring no DNA transfection and therefore would be useful if distribution data were to be gathered from the original primary cultures used to generate the DAS.

Plasmid construct. A eukaryotic expression plasmid containing a coding sequence for a green fluorescent protein – DAS chimera is prepared using GFP

[illegible] designated GFP-

-ASpp.

Cell preparation and transfection. HeLa cells are trypsinized and plated using DMEM containing 5% charcoal/dextran-treated fetal bovine serum (FBS) (Hyclone) and 1% penicillin-streptomycin (C-DMEM) 12-24 hours prior to transfection and incubated at 37°C and 5% CO_2 . Transfections are performed by calcium phosphate coprecipitation or with Lipofectamine (Life Technologies). For the calcium phosphate transfections, the medium is replaced, prior to transfection, with DMEM containing 5% charcoal/dextran-treated FBS. Cells are incubated with the calcium phosphate-DNA precipitate for 4-5 hours at 37°C and 5% CO_2, and washed 3-4 times with DMEM to remove the precipitate, followed by the addition of C-DMEM. Lipofectamine transfections are performed in serum-free DMEM without antibiotics according to the manufacturer's instructions. Following a 2-3 hour incubation with the DNA-liposome complexes, the medium is removed and replaced with C-DMEM. All transfected cells in 96-well microtiter plates are incubated at 33°C and 5% CO_2 for 24-48 hours prior to drug treatment. Experiments are performed with the receptor expressed transiently in HeLa cells.

Localization of expressed GFP-DASpp inside cells. To obtain cellular distribution data, nuclei of transfected cells are first labeled with 5 µg/ml Hoechst 33342 (Molecular Probes) in C-DMEM for 20 minutes at 33°C and 5% CO_2. Cells are washed once in Hank's Balanced Salt Solution (HBSS). The cells are analyzed live or they are rinsed with HBSS, fixed for 15 min with 3.7% formaldehyde in HBSS, stained with Hoechst 33342, and washed before analysis.

In a preferred embodiment, image acquisition and analysis are performed using the cell screening system of the present invention. The intracellular GFP-DASpp fluorescence signal is collected by acquiring fluorescence image pairs (GFP-DASpp and Hoechst 33342-labeled nuclei) from field cells. The image pairs obtained at each time point are used to define nuclear and cytoplasmic regions in each cell. Data demonstrating dispersed signal in the cytoplasm would be consistent with known

[illegible]

[illegible] confirmed for appropriate expression of the GFP-DASpp, as an indicator cell line, a screen of various ligands is performed using a series of steroid type ligands including, but not limited to: estrogen, progesterone, retinoids, growth factors,

androgens, and many other steroid and steroid based molecules. Image acquisition and analysis are performed using the cell screening system of the invention. The intracellular GFP-DASpp fluorescence signal is collected by acquiring fluorescence image pairs (GFP-DASpp and Hoechst 33342-labeled nuclei) from fields cells. The image pairs obtained at each time point are used to define nuclear and cytoplasmic regions in each cell. Translocation of GFP-DASpp is calculated by dividing the integrated fluorescence intensity of GFP-DASpp in the nucleus by the integrated fluorescence intensity of the chimera in the cytoplasm or as a nuclear-cytoplasmic difference of GFP fluorescence. A translocation from the cytoplasm into the nucleus indicates a ligand binding activation of the DASpp thus identifying the potential receptor class and action. Combining this data with other data obtained in a similar fashion using known inhibitors and modifiers of steroid receptors, would either validate the DASpp as a target, or more data would be generated from various sources.

Example 9 Additional Screens

Translocation between the plasma membrane and the cytoplasm:

Profilactin complex dissociation and binding of profilin to the plasma membrane. In one embodiment, a fluorescent protein biosensor of profilin membrane binding is prepared by labeling purified profilin (Federov et al.(1994), *J. Molec. Biol.* 241:480-482; Lanbrechts et al. (1995), *Eur. J. Biochem.* 230:281-286) with a probe possessing a fluorescence lifetime in the range of 2-300 ns. The labeled profilin is introduced into living indicator cells using bulk loading methodology and the indicator cells are treated with test compounds. Fluorescence anisotropy imaging microscopy (Gough and Taylor (1993), *J. Cell Biol.* 121:1095-1107) is used to measure test-compound dependent movement of the fluorescent derivative of profilin between the cytoplasm and membrane for a period of time after treatment ranging from 0.1 s to 10 h.

[illegible] permeabilized. The indicator cell plasma membrane, cytoplasm, and nucleus are all labeled with distinctly colored markers followed by immunolocalization of Rho protein (Self et al. (1995), *Methods in Enzymology* 256:3-10, Tanaka et al. (1995)

Methods in Enzymology 256:41-49) with antibodies labeled with a fourth color. Each of the four labels is imaged separately using the cell screening system, and the images used to calculate the amount of inhibition or activation of translocation effected by the test compound. To do this calculation, the images of the probes used to mark the plasma membrane and cytoplasm are used to mask the image of the immunological probe marking the location of intracellular Rho protein. The integrated brightness per unit area under each mask is used to form a translocation quotient by dividing the plasma membrane integrated brightness/area by the cytoplasmic integrated brightness/area. By comparing the translocation quotient values from control and experimental wells, the percent translocation is calculated for each potential lead compound.

β-Arrestin translocation to the plasma membrane upon G-protein receptor activation.

In another embodiment of a cytoplasm to membrane translocation high-content screen, the translocation of β-arrestin protein from the cytoplasm to the plasma membrane is measured in response to cell treatment. To measure the translocation, living indicator cells containing luminescent domain markers are treated with test compounds and the movement of the β-arrestin marker is measured in time and space using the cell screening system of the present invention. In a preferred embodiment, the indicator cells contain luminescent markers consisting of a green fluorescent protein β-arrestin (GFP-β-arrestin) protein chimera (Barak et al. (1997), *J. Biol. Chem.* 272:27497-27500; Daaka et al. (1998), *J. Biol. Chem.* 273:685-688) that is expressed by the indicator cells through the use of transient or stable cell transfection and other reporters used to mark cytoplasmic and membrane domains. When the indicator cells are in the resting state, the domain marker molecules partition predominately in the plasma membrane or in the cytoplasm. In the high-content screen, these markers are used to delineate the cell cytoplasm and plasma membrane in distinct channels of

[illegible] ranging from 0.1 s to 10 h. In a preferred embodiment, the time scale is 1 h. Each image is analyzed by a method that quantifies the movement of the GFP-β-arrestin

protein chimera between the plasma membrane and the cytoplasm. To do this calculation, the images of the probes used to mark the plasma membrane and cytoplasm are used to mask the image of the GFP-β-arrestin probe marking the location of intracellular GFP-β-arrestin protein. The integrated brightness per unit area under each mask is used to form a translocation quotient by dividing the plasma membrane integrated brightness/area by the cytoplasmic integrated brightness/area. By comparing the translocation quotient values from control and experimental wells, the percent translocation is calculated for each potential lead compound. The output of the high-content screen relates quantitative data describing the magnitude of the translocation within a large number of individual cells that have been treated with test compounds of interest.

Translocation between the endoplasmic reticulum and the Golgi:

In one embodiment of an endoplasmic reticulum to Golgi translocation high-content screen, the translocation of a VSVG protein from the ts045 mutant strain of vesicular stomatitis virus (Ellenberg et al. (1997), *J. Cell Biol.* 138:1193-1206; Presley et al. (1997) *Nature* 389:81-85) from the endoplasmic reticulum to the Golgi domain is measured in response to cell treatment. To measure the translocation, indicator cells containing luminescent reporters are treated with test compounds and the movement of the reporters is measured in space and time using the cell screening system of the present invention. The indicator cells contain luminescent reporters consisting of a GFP-VSVG protein chimera that is expressed by the indicator cell through the use of transient or stable cell transfection and other domain markers used to measure the localization of the endoplasmic reticulum and Golgi domains. When the indicator cells are in their resting state at 40°C, the GFP-VSVG protein chimera molecules are partitioned predominately in the endoplasmic reticulum. In this high-content screen, domain markers of distinct colors used to delineate the endoplasmic reticulum and the Golgi domains in distinct channels of fluorescence. When the indicator cells are treated with a test compound and the temperature [illegible]

[illegible] with a time scale ranging from [illegible] each image is analyzed by a method that quantifies the movement of the GFP-VSVG protein chimera between the endoplasmic reticulum and the Golgi domains. To do this calculation, the images of

the probes used to mark the endoplasmic reticulum and the Golgi domains are used to mask the image of the GFP-VSVG probe marking the location of intracellular GFP-VSVG protein. The integrated brightness per unit area under each mask is used to form a translocation quotient by dividing the endoplasmic reticulum integrated brightness/area by the Golgi integrated brightness/area. By comparing the translocation quotient values from control and experimental wells, the percent translocation is calculated for each potential lead compound. The output of the high-content screen relates quantitative data describing the magnitude of the translocation within a large number of individual cells that have been treated with test compounds of interest at final concentrations ranging from 10^{-12} M to 10^{-3} M for a period ranging from 1 min to 10 h.

Induction and inhibition of organellar function:

Intracellular microtubule stability.

In another aspect of the invention, an automated method for identifying compounds that modify microtubule structure is provided. In this embodiment, indicator cells are treated with test compounds and the distribution of luminescent microtubule-labeling molecules is measured in space and time using a cell screening system, such as the one disclosed above. The luminescent microtubule-labeling molecules may be expressed by or added to the cells either before, together with, or after contacting the cells with a test compound.

In one embodiment of this aspect of the invention, living cells express a luminescently labeled protein biosensor of microtubule dynamics, comprising a protein that labels microtubules fused to a luminescent protein. Appropriate microtubule-labeling proteins for this aspect of the invention include, but are not limited to α and β tubulin isoforms, and MAP4. Preferred embodiments of the luminescent protein include, but are not limited to green fluorescent protein (GFP) and GFP mutants. In a

limited to, α-tubulin, β-tubulin, or microtubule-associated protein 4 (MAP4). The approach outlined here enables those skilled in the art to make live cell measurements

to determine the effect of lead compounds on tubulin activity and microtubule stability *in vivo*.

In a most preferred embodiment, MAP4 is fused to a modified version of the *Aequorea victoria* green fluorescent protein (GFP). A DNA construct has been made which consists of a fusion between the EGFP coding sequence (available from Clontech) and the coding sequence for mouse MAP4. (Olson et al., (1995), J. Cell Biol. 130(3): 639-650). MAP4 is a ubiquitous microtubule-associated protein that is known to interact with microtubules in interphase as well as mitotic cells (Olmsted and Murofushi, (1993), MAP4. In "Guidebook to the Cytoskeleton and Motor Proteins." Oxford University Press. T. Kreis and R. Vale, eds.) Its localization, then, can serve as an indicator of the localization, organization, and integrity of microtubules in living (or fixed) cells at all stages of the cell cycle for cell-based HCS assays. While MAP2 and tau (microtubule associated proteins expressed specifically in neuronal cells) have been used to form GFP chimeras (Kaech *et al.*, (1996) Neuron. 17: 1189-1199; Hall *et al.*, (1997), Proc. Nat. Acad. Sci. 94: 4733-4738) their restricted cell type distribution and the tendency of these proteins to bundle microtubules when overexpressed make these proteins less desirable as molecular reagents for analysis in live cells originating from varied tissues and organs. Moderate overexpression of GFP-MAP4 does not disrupt microtubule function or integrity (Olson et al., 1995). Similar constructs can be made using β-tubulin or α-tubulin via standard techniques in the art. These chimeras will provide a means to observe and analyze microtubule activity in living cells during all stages of the cell cycle.

In another embodiment, the luminescently labeled protein biosensor of microtubule dynamics is expressed, isolated, and added to the cells to be analyzed via bulk loading techniques, such as microinjection, scrape loading, and impact-mediated loading. In this embodiment, there is not an issue of overexpression within the cell, and thus α and β tubulin isoforms, MAP4, MAP2 and/or tau can all be used.

[illegible] detects the protein biosensor, endogenous levels of a protein antigen, or both. In this embodiment, a luminescent label that detects α and β tubulin isoforms, MAP4, MAP2 and/or tau, can be used.

A variety of GFP mutants are available, all of which would be effective in this invention, including, but not limited to, GFP mutants which are commercially available (Clontech, California).

The MAP4 construct has been introduced into several mammalian cell lines (BHK-21, Swiss 3T3, HeLa, HEK 293, LLCPK) and the organization and localization of tubulin has been visualized in live cells by virtue of the GFP fluorescence as an indicator of MAP4 localization. The construct can be expressed transiently or stable cell lines can be prepared by standard methods. Stable HeLa cell lines expressing the EGFP-MAP4 chimera have been obtained, indicating that expression of the chimera is not toxic and does not interfere with mitosis.

Possible selectable markers for establishment and maintenance of stable cell lines include, but are not limited to the neomycin resistance gene, hygromycin resistance gene, zeocin resistance gene, puromycin resistance gene, bleomycin resistance gene, and blastacidin resistance gene.

The utility of this method for the monitoring of microtubule assembly, disassembly, and rearrangement has been demonstrated by treatment of transiently and stably transfected cells with microtubule drugs such as paclitaxel, nocodazole, vincristine, or vinblastine.

The present method provides high-content and combined high throughput-high content cell-based screens for anti-microtubule drugs, particularly as one parameter in a multi-parametric cancer target screen. The EGFP-MAP4 construct used herein can also be used as one of the components of a high-content screen that measures multiple signaling pathways or physiological events. In a preferred embodiment, a combined high throughput and high content screen is employed, wherein multiple cells in each of the locations containing cells are analyzed in a high throughput mode, and only a subset of the locations containing cells are analyzed in a high content mode. The high throughput screen can be any screen that would be useful to identify those locations

expression of a reporter gene, those undergoing calcium changes, and those undergoing pH changes.

In addition to drug screening applications, the present invention may be applied to clinical diagnostics, the detection of chemical and biological warfare weapons, and the basic research market since fundamental cell processes, such as cell division and motility, are highly dependent upon microtubule dynamics.

5

Image Acquisition and Analysis

Image data can be obtained from either fixed or living indicator cells. To extract morphometric data from each of the images obtained the following method of analysis is used:

10 1. Threshold each nucleus and cytoplasmic image to produce a mask that has value = 0 for each pixel outside a nucleus or cell boundary.
2. Overlay the mask on the original image, detect each object in the field (*i.e.*, nucleus or cell), and calculate its size, shape, and integrated intensity.
3. Overlay the whole cell mask obtained above on the corresponding luminescent 15 microtubule image and apply one or more of the following set of classifiers to determine the micrtotubule morphology and the effect of drugs on microtubule morphology.

Microtubule morphology is defined using a set of classifiers to quantify aspects of microtubule shape, size, aggregation state, and polymerization state. These 20 classifiers can be based on approaches that include co-occurrence matrices, texture measurements, spectral methods, structural methods, wavelet transforms, statistical methods, or combinations thereof. Examples of such classifiers are as follows:

1. A classifier to quantify microtubule length and width using edge detection methods such as that discussed in Kolega et al. ((1993). *BioImaging* 1:136-25 150), which discloses a non-automated method to determine edge strength in individual cells), to calculate the total edge strength within each cell. To normalize for cell size, the total edge strength can be divided by the cell area to give a "microtubule morphology" value. Large microtubule morphology values are associated with strong edge strength values and are therefore maximal in cells containing distinct microtubule 30 structures. Likewise, small microtubule morphology values are associated with weak

[illegible] microtubule [illegible] paclitaxel [illegible] the microtubule depolymerizing drug nocodazole (10 μg/ml).

35

2. A classifier to quantify microtubule aggregation into punctate spots or foci using methodology from the receptor internalization methods discussed supra

3. A classifier to quantify microtubule depolymerization using a measure of image texture.

4. A classifier to quantify apparent interconnectivity, or branching (or both), of the microtubules.

5. Measurement of the kinetics of microtubule reorganization using the above classifiers on a time series of images of cells treated with test compounds.

In a further aspect, kits are provided for analyzing microtubule stability, comprising an expression vector comprising a nucleic acid that encodes a microtubule labeling protein and instructions for using the expression vector for carrying out the methods described above. In a preferred embodiment, the expression vector further comprises a nucleic acid that encodes a luminescent protein, wherein the microtubule binding protein and the luminescent protein thereof are expressed as a fusion protein. Alternatively, the kit may contain an antibody that specifically binds to the microtubule-labeling protein. In a further embodiment, the kit includes cells that express the microtubule labeling protein. In a preferred embodiment, the cells are transfected with the expression vector. In another preferred embodiment, the kits further contain a compound that is known to disrupt microtubule structure, including but not limited to curacin, nocodazole, vincristine, or vinblastine. In another preferred embodiment, the kits further comprise a compound that is known to stabilize microtubule structure, including but not limited to taxol (paclitaxel), and discodermolide.

In another aspect, the present invention comprises a machine readable storage medium comprising a program containing a set of instructions for causing a cell screening system to execute the disclosed methods for analyzing microtubule stability, wherein the cell screening system comprises an optical system with a stage adapted for holding a plate containing cells, a digital camera, a means for directing fluorescence or luminescence emitted from

High-content screens involving the functional localization of macromolecules

Within this class of high-content screen, the functional localization of macromolecules in response to external stimuli is measured within living cells.

Glycolytic enzyme activity regulation. In a preferred embodiment of a cellular enzyme activity high-content screen, the activity of key glycolytic regulatory enzymes are measured in treated cells. To measure enzyme activity, indicator cells containing luminescent labeling reagents are treated with test compounds and the activity of the reporters is measured in space and time using cell screening system of the present invention.

In one embodiment, the reporter of intracellular enzyme activity is fructose-6-phosphate, 2-kinase/fructose-2,6-bisphosphatase (PFK-2), a regulatory enzyme whose phosphorylation state indicates intracellular carbohydrate anabolism or catabolism (Deprez et al. (1997) *J. Biol. Chem.* 272:17269-17275; Kealer et al. (1996) *FEBS Letters* 395:225-227; Lee et al. (1996), *Biochemistry* 35:6010-6019). The indicator cells contain luminescent reporters consisting of a fluorescent protein biosensor of PFK-2 phosphorylation. The fluorescent protein biosensor is constructed by introducing an environmentally sensitive fluorescent dye near to the known phosphorylation site of the enzyme (Deprez et al. (1997), *supra*; Giuliano et al. (1995), *supra*). The dye can be of the ketocyanine class (Kessler and Wolfbeis (1991), *Spectrochimica Acta* 47A:187-192) or any class that contains a protein reactive moiety and a fluorochrome whose excitation or emission spectrum is sensitive to solution polarity. The fluorescent protein biosensor is introduced into the indicator cells using bulk loading methodology.

Living indicator cells are treated with test compounds, at final concentrations ranging from 10^{-12} M to 10^{-3} M for times ranging from 0.1 s to 10 h. In a preferred embodiment, ratio image data are obtained from living treated indicator cells by collecting a spectral pair of fluorescence images at each time point. To extract

[illegible] then used to calculate the fractional phosphorylation of PFK-2. At small fractional values of phosphorylation, PFK-2 stimulates carbohydrate catabolism.

At high fractional values of phosphorylation, PFK-2 stimulates carbohydrate anabolism.

Protein kinase A activity and localization of subunits. In another embodiment of a high-content screen, both the domain localization and activity of protein kinase A (PKA) within indicator cells are measured in response to treatment with test compounds.

The indicator cells contain luminescent reporters including a fluorescent protein biosensor of PKA activation. The fluorescent protein biosensor is constructed by introducing an environmentally sensitive fluorescent dye into the catalytic subunit of PKA near the site known to interact with the regulatory subunit of PKA (Harootunian et al. (1993), *Mol. Biol. of the Cell* 4:993-1002; Johnson et al. (1996), *Cell* 85:149-158; Giuliano et al. (1995), *supra*). The dye can be of the ketocyanine class (Kessler, and Wolfbeis (1991), *Spectrochimica Acta* 47A:187-192) or any class that contains a protein reactive moiety and a fluorochrome whose excitation or emission spectrum is sensitive to solution polarity. The fluorescent protein biosensor of PKA activation is introduced into the indicator cells using bulk loading methodology.

In one embodiment, living indicator cells are treated with test compounds, at final concentrations ranging from 10^{-12} M to 10^{-3} M for times ranging from 0.1 s to 10 h. In a preferred embodiment, ratio image data are obtained from living treated indicator cells. To extract biosensor data from each time point, a ratio is made between each pair of images, and each pixel value is then used to calculate the fractional activation of PKA (*e.g.*, separation of the catalytic and regulatory subunits after cAMP binding). At high fractional values of activity, PFK-2 stimulates biochemical cascades within the living cell.

To measure the translocation of the catalytic subunit of PKA, indicator cells containing luminescent reporters are treated with test compounds and the movement of

measure the localization of the cytoplasmic and nuclear domains. When the indicator cells are treated with a test compounds, the dynamic redistribution of a PKA fluorescent protein biosensor is recorded intracellularly as a series of images over a

time scale ranging from 0.1 s to 10 h. Each image is analyzed by a method that quantifies the movement of the PKA between the cytoplasmic and nuclear domains. To do this calculation, the images of the probes used to mark the cytoplasmic and nuclear domains are used to mask the image of the PKA fluorescent protein biosensor. The integrated brightness per unit area under each mask is used to form a translocation quotient by dividing the cytoplasmic integrated brightness/area by the nuclear integrated brightness/area. By comparing the translocation quotient values from control and experimental wells, the percent translocation is calculated for each potential lead compound. The output of the high-content screen relates quantitative data describing the magnitude of the translocation within a large number of individual cells that have been treated with test compound in the concentration range of 10^{-12} M to 10^{-3} M.

High-content screens involving the induction or inhibition of gene expression

RNA-based fluorescent biosensors

Cytoskeletal protein transcription and message localization. Regulation of the general classes of cell physiological responses including cell-substrate adhesion, cell-cell adhesion, signal transduction, cell-cycle events, intermediary and signaling molecule metabolism, cell locomotion, cell-cell communication, and cell death can involve the alteration of gene expression. High-content screens can also be designed to measure this class of physiological response.

In one embodiment, the reporter of intracellular gene expression is an oligonucleotide that can hybridize with the target mRNA and alter its fluorescence signal. In a preferred embodiment, the oligonucleotide is a molecular beacon (Tyagi and Kramer (1996) *Nat. Biotechnol.* 14:303-308), a luminescence-based reagent whose fluorescence signal is dependent on intermolecular and intramolecular interactions. The fluorescent biosensor is constructed by introducing a fluorescence energy transfer pair of fluorescent dyes

whose excitation and emission spectra overlap sufficiently to provide fluorescence energy transfer between the dyes in the resting state, including, but not limited to, fluorescein and rhodamine (Molecular Probes, Inc.). In a preferred embodiment, a

portion of the message coding for β-actin (Kislauskis et al. (1994), *J. Cell Biol.* 127:441-451; McCann et al. (1997), *Proc. Natl. Acad. Sci.* 94:5679-5684; Sutoh (1982), *Biochemistry* 21:3654-3661) is inserted into the loop region of a hairpin-shaped oligonucleotide with the ends tethered together due to intramolecular hybridization. At each end of the biosensor a fluorescence donor (fluorescein) and a fluorescence acceptor (rhodamine) are covalently bound. In the tethered state, the fluorescence energy transfer is maximal and therefore indicative of an unhybridized molecule. When hybridized with the mRNA coding for β-actin, the tether is broken and energy transfer is lost. The complete fluorescent biosensor is introduced into the indicator cells using bulk loading methodology.

In one embodiment, living indicator cells are treated with test compounds, at final concentrations ranging from 10^{-12} M to 10^{-3} M for times ranging from 0.1 s to 10 h. In a preferred embodiment, ratio image data are obtained from living treated indicator cells. To extract morphometric data from each time point, a ratio is made between each pair of images, and each pixel value is then used to calculate the fractional hybridization of the labeled nucleotide. At small fractional values of hybridization little expression of β-actin is indicated. At high fractional values of hybridization, maximal expression of β-actin is indicated. Furthermore, the distribution of hybridized molecules within the cytoplasm of the indicator cells is also a measure of the physiological response of the indicator cells.

Cell surface binding of a ligand

Labeled insulin binding to its cell surface receptor in living cells. Cells whose plasma membrane domain has been labeled with a labeling reagent of a particular color are incubated with a solution containing insulin molecules (Lee et al. (1997), *Biochemistry* 36:2701-2708; Martinez-Zaguilan et al. (1996), *Am. J. Physiol.* 270:C1438-C1446) that are labeled with a luminescent probe of a different color for an

insulin on the plasma membrane is measured. To do this, the cell membrane image is used as a mask for the insulin image. The integrated intensity from the masked insulin image is compared to a set of images containing known amounts of labeled insulin

The amount of insulin bound to the cell is determined from the standards and used in conjunction with the total concentration of insulin incubated with the cell to calculate a dissociation constant or insulin to its cell surface receptor.

Labeling of cellular compartments

Whole cell labeling

Whole cell labeling is accomplished by labeling cellular components such that dynamics of cell shape and motility of the cell can be measured over time by analyzing fluorescence images of cells.

In one embodiment, small reactive fluorescent molecules are introduced into living cells. These membrane-permeant molecules both diffuse through and react with protein components in the plasma membrane. Dye molecules react with intracellular molecules to both increase the fluorescence signal emitted from each molecule and to entrap the fluorescent dye within living cells. These molecules include reactive chloromethyl derivatives of aminocoumarins, hydroxycoumarins, eosin diacetate, fluorescein diacetate, some Bodipy dye derivatives, and tetramethylrhodamine. The reactivity of these dyes toward macromolecules includes free primary amino groups and free sulfhydryl groups.

In another embodiment, the cell surface is labeled by allowing the cell to interact with fluorescently labeled antibodies or lectins (Sigma Chemical Company, St. Louis, MO) that react specifically with molecules on the cell surface. Cell surface protein chimeras expressed by the cell of interest that contain a green fluorescent protein, or mutant thereof, component can also be used to fluorescently label the entire cell surface. Once the entire cell is labeled, images of the entire cell or cell array can become a parameter in high content screens, involving the measurement of cell shape, motility, size, and growth and division.

[illegible] plasma membrane employs some of the same methodology described above for labeling the entire cells. Luminescent molecules that label the entire cell surface act to delineate the plasma membrane.

In a second embodiment subdomains of the plasma membrane, the extracellular surface, the lipid bilayer, and the intracellular surface can be labeled separately and used as components of high content screens. In the first embodiment, the extracellular surface is labeled using a brief treatment with a reactive fluorescent molecule such as the succinimidyl ester or iodoacetamde derivatives of fluorescent dyes such as the fluoresceins, rhodamines, cyanines, and Bodipys.

In a third embodiment, the extracellular surface is labeled using fluorescently labeled macromolecules with a high affinity for cell surface molecules. These include fluorescently labeled lectins such as the fluorescein, rhodamine, and cyanine derivatives of lectins derived from jack bean (Con A), red kidney bean (erythroagglutinin PHA-E), or wheat germ.

In a fourth embodiment, fluorescently labeled antibodies with a high affinity for cell surface components are used to label the extracellular region of the plasma membrane. Extracellular regions of cell surface receptors and ion channels are examples of proteins that can be labeled with antibodies.

In a fifth embodiment, the lipid bilayer of the plasma membrane is labeled with fluorescent molecules. These molecules include fluorescent dyes attached to long chain hydrophobic molecules that interact strongly with the hydrophobic region in the center of the plasma membrane lipid bilayer. Examples of these dyes include the PKH series of dyes (U.S. 4,783,401, 4,762701, and 4,859,584; available commercially from Sigma Chemical Company, St. Loius, MO), fluorescent phospholipids such as nitrobenzoxadiazole glycerophosphoethanolamine and fluorescein-derivatized dihexadecanoylglycerophosphoetha-nolamine, fluorescent fatty acids such as 5-butyl-4,4-difluoro-4-bora-3a,4a-diaza-s-indacene-3-nonanoic acid and 1-pyrenedecanoic acid (Molecular Probes, Inc.), fluorescent sterols including cholesteryl 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-dodecanoate and cholesteryl 1-pyrenehexanoate, and fluorescently labeled proteins that interact specifically with lipid

another embodiment, the intracellular component of the plasma membrane is labeled with fluorescent molecules. Examples of these molecules are the intracellular components of the trimeric G-protein receptor, adenylyl cyclase, and ionic transpor

proteins. These molecules can be labeled as a result of tight binding to a fluorescently labeled specific antibody or by the incorporation of a fluorescent protein chimera that is comprised of a membrane-associated protein and the green fluorescent protein, and mutants thereof.

Endosome fluorescence labeling

In one embodiment, ligands that are transported into cells by receptor-mediated endocytosis are used to trace the dynamics of endosomal organelles. Examples of labeled ligands include Bodipy FL-labeled low density lipoprotein complexes, tetramethylrhodamine transferrin analogs, and fluorescently labeled epidermal growth factor (Molecular Probes, Inc.)

In a second embodiment, fluorescently labeled primary or secondary antibodies (Sigma Chemical Co. St. Louis, MO; Molecular Probes, Inc. Eugene, OR; Caltag Antibody Co.) that specifically label endosomal ligands are used to mark the endosomal compartment in cells.

In a third embodiment, endosomes are fluorescently labeled in cells expressing protein chimeras formed by fusing a green fluorescent protein, or mutants thereof, with a receptor whose internalization labels endosomes. Chimeras of the EGF, transferrin, and low density lipoprotein receptors are examples of these molecules.

Lysosome labeling

In one embodiment, membrane permeant lysosome-specific luminescent reagents are used to label the lysosomal compartment of living and fixed cells. These reagents include the luminescent molecules neutral red, N-(3-((2,4-dinitrophenyl)amino)propyl)-N-(3-aminopropyl)methylamine, and the LysoTracker probes which report intralysosomal pH as well as the dynamic distribution of lysosomes (Molecular Probes, Inc.)

In a second embodiment, [illegible] lysosomal components that are localized in specific lysosomal domains. Examples of these components are the degradative enzymes involved in cholesterol ester hydrolysis,

membrane protein proteases, and nucleases as well as the ATP-driven lysosomal proton pump.

In a third embodiment, protein chimeras consisting of a lysosomal protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label the lysosomal domain. Examples of these components are the degradative enzymes involved in cholesterol ester hydrolysis, membrane protein proteases, and nucleases as well as the ATP-driven lysosomal proton pump.

Cytoplasmic fluorescence labeling

In one embodiment, cell permeant fluorescent dyes (Molecular Probes, Inc.) with a reactive group are reacted with living cells. Reactive dyes including monobromobimane, 5-chloromethylfluorescein diacetate, carboxy fluorescein diacetate succinimidyl ester, and chloromethyl tetramethylrhodamine are examples of cell permeant fluorescent dyes that are used for long term labeling of the cytoplasm of cells.

In a second embodiment, polar tracer molecules such as Lucifer yellow and cascade blue-based fluorescent dyes (Molecular Probes, Inc.) are introduced into cells using bulk loading methods and are also used for cytoplasmic labeling.

In a third embodiment, antibodies against cytoplasmic components (Sigma Chemical Co.; Molecular Probes, Inc.; Caltag Antibody Co.) are used to fluorescently label the cytoplasm. Examples of cytoplasmic antigens are many of the enzymes involved in intermediary metabolism. Enolase, phosphofructokinase, and acetyl-CoA dehydrogenase are examples of uniformly distributed cytoplasmic antigens.

In a fourth embodiment, protein chimeras consisting of a cytoplasmic protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label the cytoplasm. Fluorescent chimeras of uniformly distributed proteins are used to label the entire cytoplasmic domain. Examples of these proteins

In a fifth embodiment, antibodies against cytoplasmic antigens (Sigma Chemical Co.; Molecular Probes, Inc.; Caltag Antibody Co.) are used to label cytoplasmic components that are localized in specific cytoplasmic subdomains

Examples of these components are the cytoskeletal proteins actin, tubulin, and cytokeratin. A population of these proteins within cells is assembled into discrete structures, which in this case, are fibrous. Fluorescence labeling of these proteins with antibody-based reagents therefore labels a specific sub-domain of the cytoplasm.

In a sixth embodiment, non-antibody-based fluorescently labeled molecules that interact strongly with cytoplasmic proteins are used to label specific cytoplasmic components. One example is a fluorescent analog of the enzyme DNAse I (Molecular Probes, Inc.) Fluorescent analogs of this enzyme bind tightly and specifically to cytoplasmic actin, thus labeling a sub-domain of the cytoplasm. In another example, fluorescent analogs of the mushroom toxin phalloidin or the drug paclitaxel (Molecular Probes, Inc.) are used to label components of the actin- and microtubule-cytoskeletons, respectively.

In a seventh embodiment, protein chimeras consisting of a cytoplasmic protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label specific domains of the cytoplasm. Fluorescent chimeras of highly localized proteins are used to label cytoplasmic sub-domains. Examples of these proteins are many of the proteins involved in regulating the cytoskeleton. They include the structural proteins actin, tubulin, and cytokeratin as well as the regulatory proteins microtubule associated protein 4 and α-actinin.

Nuclear labeling

In one embodiment, membrane permeant nucleic-acid-specific luminescent reagents (Molecular Probes, Inc.) are used to label the nucleus of living and fixed cells. These reagents include cyanine-based dyes (*e.g.*, TOTO®, YOYO®, and BOBO™), phenanthidines and acridines (*e.g.*, ethidium bromide, propidium iodide, and acridine orange), indoles and imidazoles (*e.g.*, Hoechst 33258, Hoechst 33342, and 4',6-diamidino-2-phenylindole), and other similar reagents (*e.g.*, 7-aminoactinomycin D,

Molecular Probes, Inc., Caltag Antibody Co.) are used to label nuclear components that are localized in specific nuclear domains. Examples of these components are the macromolecules involved in maintaining DNA structure and

function. DNA, RNA, histones, DNA polymerase, RNA polymerase, lamins, and nuclear variants of cytoplasmic proteins such as actin are examples of nuclear antigens.

In a third embodiment, protein chimeras consisting of a nuclear protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label the nuclear domain. Examples of these proteins are many of the proteins involved in maintaining DNA structure and function. Histones, DNA polymerase, RNA polymerase, lamins, and nuclear variants of cytoplasmic proteins such as actin are examples of nuclear proteins.

Mitochondrial labeling

In one embodiment, membrane permeant mitochondrial-specific luminescent reagents (Molecular Probes, Inc.) are used to label the mitochondria of living and fixed cells. These reagents include rhodamine 123, tetramethyl rosamine, JC-1, and the MitoTracker reactive dyes.

In a second embodiment, antibodies against mitochondrial antigens (Sigma Chemical Co.; Molecular Probes, Inc.; Caltag Antibody Co.) are used to label mitochondrial components that are localized in specific mitochondrial domains. Examples of these components are the macromolecules involved in maintaining mitochondrial DNA structure and function. DNA, RNA, histones, DNA polymerase, RNA polymerase, and mitochondrial variants of cytoplasmic macromolecules such as mitochondrial tRNA and rRNA are examples mitochondrial antigens. Other examples of mitochondrial antigens are the components of the oxidative phosphorylation system found in the mitochondria (*e.g.*, cytochrome c, cytochrome c oxidase, and succinate dehydrogenase).

In a third embodiment, protein chimeras consisting of a mitochondrial protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label the mitochondrial domain. Examples of

merase, and the components of the oxidative phosphorylation system found in the mitochondria (*e.g.*, cytochrome c, cytochrome c oxidase, and succinate dehydrogenase).

Endoplasmic reticulum labeling

In one embodiment, membrane permeant endoplasmic reticulum-specific luminescent reagents (Molecular Probes, Inc.) are used to label the endoplasmic reticulum of living and fixed cells. These reagents include short chain carbocyanine dyes (*e.g.*, $DiOC_6$ and $DiOC_3$), long chain carbocyanine dyes (*e.g.*, $DiIC_{16}$ and $DiIC_{18}$), and luminescently labeled lectins such as concanavalin A.

In a second embodiment, antibodies against endoplasmic reticulum antigens (Sigma Chemical Co.; Molecular Probes, Inc.; Caltag Antibody Co.) are used to label endoplasmic reticulum components that are localized in specific endoplasmic reticulum domains. Examples of these components are the macromolecules involved in the fatty acid elongation systems, glucose-6-phosphatase, and HMG CoA-reductase.

In a third embodiment, protein chimeras consisting of a endoplasmic reticulum protein genetically fused to an intrinsically luminescent protein such as the green fluorescent protein, or mutants thereof, are used to label the endoplasmic reticulum domain. Examples of these components are the macromolecules involved in the fatty acid elongation systems, glucose-6-phosphatase, and HMG CoA-reductase.

Golgi labeling

In one embodiment, membrane permeant Golgi-specific luminescent reagents (Molecular Probes, Inc.) are used to label the Golgi of living and fixed cells. These reagents include luminescently labeled macromolecules such as wheat germ agglutinin and Brefeldin A as well as luminescently labeled ceramide.

In a second embodiment, antibodies against Golgi antigens (Sigma Chemical Co.; Molecular Probes, Inc.; Caltag Antibody Co.) are used to label Golgi components that are localized in specific Golgi domains. Examples of these components are N-acetylglucosamine phosphotransferase, Golgi-specific phosphodiesterase, and mannose-6-phosphate receptor protein.

In a third embodiment, protein chimeras consisting of a Golgi protein

components are N-acetylglucosamine phosphotransferase, Golgi-specific phosphodiesterase, and mannose-6-phosphate receptor protein.

While many of the examples presented involve the measurement of single cellular processes, this is again is intended for purposes of illustration only. Multiple parameter high-content screens can be produced by combining several single parameter screens into a multiparameter high-content screen or by adding cellular parameters to any existing high-content screen. Furthermore, while each example is described as being based on either live or fixed cells, each high-content screen can be designed to be used with both live and fixed cells.

Those skilled in the art will recognize a wide variety of distinct screens that can be developed based on the disclosure provided herein. There is a large and growing list of known biochemical and molecular processes in cells that involve translocations or reorganizations of specific components within cells. The signaling pathway from the cell surface to target sites within the cell involves the translocation of plasma membrane-associated proteins to the cytoplasm. For example, it is known that one of the src family of protein tyrosine kinases, pp60c-src (Walker et al (1993), *J. Biol. Chem.* 268:19552-19558) translocates from the plasma membrane to the cytoplasm upon stimulation of fibroblasts with platelet-derived growth factor (PDGF). Additionally, the targets for screening can themselves be converted into fluorescence-based reagents that report molecular changes including ligand-binding and post-translocational modifications.

Protease Biosensors

(1) Background

As used herein, the following terms are defined as follows:

- Reactant – the parent biosensor that interacts with the proteolytic enzyme.
- Product – the signal-containing proteolytic fragment(s) generated by the interaction of the reactant with the enzyme.
- Reactant Target Sequence – an amino acid sequence that imparts a restriction on the

[illegible] distribution of the signal-containing product(s) of the targeted enzymatic reaction to a particular subcellular domain of the cell. If the product is initially localized within a membrane bound compartment, then the Product Target

Sequence must incorporate the ability to export the product out of the membrane-bound compartment. A bi-functional sequence can be used, which first moves the product out of the membrane-bound compartment, and then targets the product to the final compartment. In general, the same amino acid sequences can act as either or both reactant target sequences and product target sequences. Exceptions to this include amino acid sequences which target the nuclear envelope, Golgi apparatus, endoplasmic reticuulum, and which are involved in farnesylation, which are more suitable as reactant target sequences.

- Protease Recognition Site – an amino acid sequence that imparts specificity by mimicking the substrate, providing a specific binding and cleavage site for a protease. Although typically a short sequence of amino acids representing the minimal cleavage site for a protease (e.g. DEVD for caspase-3, Villa, P., S.H. Kaufmann, and W.C. Earnshaw. 1997. Caspases and caspase inhibitors. *Trends Biochem Sci.* 22:388-93), greater specificity may be established by using a longer sequence from an established substrate.
- Compartment – any cellular sub-structure or macromolecular component of the cell, whether it is made of protein, lipid, carbohydrate, or nucleic acid. It could be a macromolecular assembly or an organelle (a membrane delimited cellular component). Compartments include, but are not limited to, cytoplasm, nucleus, nucleolus, inner and outer surface of nuclear envelope, cytoskeleton, peroxisome, endosome, lysosome, inner leaflet of plasma membrane, outer leaflet of plasma membrane, outer leaflet of mitochondrial membrane, inner leaflet of mitochondrial membrane, Golgi, endoplasmic reticulum, or extracellular space.

Signal – an amino acid sequence that can be detected. This includes, but is not limited to inherently fluorescent proteins (e.g. Green Fluorescent Protein), cofactor-requiring fluorescent or luminescent proteins (e.g. phycobiliproteins or luciferases), and epitopes recognizable by specific antibodies or other specific natural or

[illegible] included are site-specifically labeled proteins that contain a luminescent dye. Methodology for site-specific labeling of proteins includes, but is not limited to, engineered dye-reactive amino acids (Post, et al., *J. Biol. Chem.* 269:12880-12887

(1994)), enzyme-based incorporation of luminescent substrates into proteins (Buckler, et al., *Analyt. Biochem.* 209:20-31 (1993); Takashi, *Biochemistry.* 27:938-943 (1988)), and the incorporation of unnatural labeled amino acids into proteins (Noren, et al., *Science.* 244:182-188 (1989)).

- Detection – a means for recording the presence, position, or amount of the signal. The approach may be direct, if the signal is inherently fluorescent, or indirect, if, for example, the signal is an epitope that must be subsequently detected with a labeled antibody. Modes of detection include, but are not limited to, the spatial position of fluorescence, luminescence, or phosphorescence: (1) intensity; (2) polarization; (3) lifetime; (4) wavelength; (5) energy transfer; and (6) recovery after photobleaching.

The basic principle of the protease biosensors of the present invention is to spatially separate the reactants from the products generated during a proteolytic reaction. The separation of products from reactants occurs upon proteolytic cleavage of the protease recognition site within the biosensor, allowing the products to bind to, diffuse into, or be imported into compartments of the cell different from those of the reactant. This spatial separation provides a means of quantitating a proteolytic process directly in living or fixed cells. Some designs of the biosensor provide a means of restricting the reactant (uncleaved biosensor) to a particular compartment by a protein sequence ("reactant target sequence") that binds to or imports the biosensor into a compartment of the cell. These compartments include, but are not limited to any cellular substructure, macromolecular cellular component, membrane-limited organelles, or the extracellular space. Given that the characteristics of the proteolytic reaction are related to product concentration divided by the reactant concentration, the spatial separation of products and reactants provides a means of uniquely quantitating products and reactants in single cells, allowing a more direct measure of proteolytic activity.

The molecular-based biosensors may be introduced into cells via transfection

eukaryotic expression system, and the purified protein introduced into the cell via a number of physical mechanisms including, but not limited to, micro-injection, scrape loading, electroporation, signal-sequence mediated loading, etc

Measurement modes may include, but are not limited to, the ratio or difference in fluorescence, luminescence, or phosphorescence: (a) intensity; (b) polarization; or (c) lifetime between reactant and product. These latter modes require appropriate spectroscopic differences between products and reactants. For example, cleaving a reactant containing a limited-mobile signal into a very small translocating component and a relatively large non-translocating component may be detected by polarization. Alternatively, significantly different emission lifetimes between reactants and products allow detection in imaging and non-imaging modes.

One example of a family of enzymes for which this biosensor can be constructed to report activity is the caspases. Caspases are a class of proteins that catalyze proteolytic cleavage of a wide variety of targets during apoptosis. Following initiation of apoptosis, the Class II "downstream" caspases are activated and are the point of no return in the pathway leading to cell death, resulting in cleavage of downstream target proteins. In specific examples, the biosensors described here were engineered to use nuclear translocation of cleaved GFP as a measurable indicator of caspase activation. Additionally, the use of specific recognition sequences that incorporate surrounding amino acids involved in secondary structure formation in naturally occurring proteins may increase the specificity and sensitivity of this class of biosensor.

Another example of a protease class for which this biosensor can be constructed to report activity is zinc metalloproteases. Two specific examples of this class are the biological toxins derived from *Clostridial* species (*C. botulinum* and *C. tetani*) and *Bacillus anthracis*. (Herreros et al. *In* The Comprehensive Sourcebook of Bacterial Protein Toxins. J.E. Alouf and J.H. Freer, Eds. 2nd edition, San Diego, Academic Press, 1999; pp 202-228.) These bacteria express and secrete zinc metalloproteases that enter eukaryotic cells and specifically cleave distinct target proteins. For example, the anthrax protease from *Bacillus anthracis* is delivered into the cytoplasm of target cells via an accessory pore forming protein

kinases [illegible] MEK1 or MEK2. [illegible] The Comprehensive Sourcebook of Bacterial Protein Toxins. J.E. Alouf and J.H. Freer, Eds. 2nd edition, San Diego, Academic Press, 1999; pp243-263.) The toxin biosensors described here take

advantage of the natural subcellular localization of these and other target proteins to achieve reactant targeting. Upon cleavage, the signal (with or without a product target sequence) is separated from the reactant to create a high-content biosensor.

One of skill in the art will recognize that the protein biosensors of this aspect of the invention can be adapted to report the activity of any member of the caspase family of proteases, as well as any other protease, by a substitution of the appropriate protease recognition site in any of the constructs (**see Figure 29B**). These biosensors can be used in high-content screens to detect in vivo activation of enzymatic activity and to identify specific activity based on cleavage of a known recognition motif. This screen can be used for both live cell and fixed end-point assays, and can be combined with additional measurements to provide a multi-parameter assay.

Thus, in another aspect the present invention provides recombinant nucleic acids encoding a protease biosensor, comprising:

a. a first nucleic acid sequence that encodes at least one detectable polypeptide signal;

b. a second nucleic acid sequence that encodes at least one protease recognition site, wherein the second nucleic acid sequence is operatively linked to the first nucleic acid sequence that encodes the at least one detectable polypeptide signal; and

c. a third nucleic acid sequence that encodes at least one reactant target sequence, wherein the third nucleic acid sequence is operatively linked to the second nucleic acid sequence that encodes the at least one protease recognition site.

In this aspect, the first and third nucleic acid sequences are separated by the second nucleic acid sequence, which encodes the protease recognition site.

In a further embodiment, the recombinant nucleic acid encoding a protease biosensor comprises a fourth nucleic acid sequence that encodes at least one product

In a further embodiment, the recombinant nucleic acid encoding a protease biosensor comprises a fifth nucleic acid sequence that encodes at least one detectable

polypeptide signal, wherein the fifth nucleic acid sequence is operatively linked to the third nucleic acid sequence that encodes the reactant target sequence.

In a preferred embodiment, the detectable polypeptide signal is selected from the group consisting of fluorescent proteins, luminescent proteins, and sequence epitopes. In a most preferred embodiment, the first nucleic acid encoding a polypeptide sequence comprises a sequence selected from the group consisting of **SEQ ID NOS: 35, 37, 39, 41, 43, 45, 47, 49, and 51.**

In another preferred embodiment, the second nucleic acid encoding a protease recognition site comprises a sequence selected from the group consisting of **SEQ ID NOS: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, and 121.** In another preferred embodiment, the third nucleic acid encoding a reactant target sequence comprises a sequence selected from the group consisting of **SEQ ID NOS: 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, and 151.**

In a most preferred embodiment, the recombinant nucleic acid encoding a protease biosensor comprises a sequence substantially similar to sequences selected from the group consisting of **SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.**

In another aspect, the present invention provides a recombinant expression vector comprising nucleic acid control sequences operatively linked to the above-described recombinant nucleic acids. In a still further aspect, the present invention provides genetically engineered host cells that have been transfected with the recombinant expression vectors of the invention.

In another aspect, the present invention provides recombinant protease biosensors comprising

a. a first domain comprising at least one detectable polypeptide signal;

[illegible] domain comprising at least one reactant target sequence,

wherein the first domain and the third domain are separated by the second domain.

Inherent in this embodiment is the concept that the reactant target sequence restricts the cellular distribution of the reactant, with redistribution of the product occurring after activation (ie: protease cleavage). This redistribution does not require a complete sequestration of products and reactants, as the product distribution can partially overlap the reactant distribution in the absence of a product targeting signal (see below).

In a preferred embodiment, the recombinant protease biosensor further comprises a fourth domain comprising at least one product target sequence, wherein the fourth domain and the first domain are operatively linked and are separated from the third domain by the second domain. In another embodiment, the recombinant protease biosensor further comprises a fifth domain comprising at least one detectable polypeptide signal, wherein the fifth domain and the third domain are operatively linked and are separated from the first domain by the second domain.

In a preferred embodiment, the detectable polypeptide signal domain (first or fifth domain) is selected from the group consisting of fluorescent proteins, luminescent proteins, and sequence epitopes. In a most preferred embodiment, the detectable polypeptide signal domain comprises a sequence selected from the group consisting of **SEQ ID NOS:36, 38, 40, 42, 44, 46, 48, 50, and 52.**

In another preferred embodiment, the second domain comprising a protease recognition site comprises a sequence selected from the group consisting of **SEQ ID NOS:54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, and 122.** In another preferred embodiment, the reactant and/or target sequence domains comprise a sequence selected from the group consisting of **SEQ ID NOS:124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, and 152.**

In a most preferred embodiment, the recombinant protease biosensor comprises

In still further embodiment, the present invention provides methods and kits for automated analysis of cells, comprising using cells that possess the protease biosensors of the invention to identify compounds that affect protease activity. The

method can be combined with the other methods of the invention in a variety of possible multi-parametric assays.

In these various embodiments, the basic protease biosensor is composed of multiple domains, including at least a first detectable polypeptide signal domain, at least one reactant target domain, and at least one protease recognition domain, wherein the detectable signal domain and the reactant target domain are separated by the protease recognition domain. Thus, the exact order of the domains in the molecule is not generally critical, so long as the protease recognition domain separates the reactant target and first detectable signal domain. For each domain, one or more one of the specified recognition sequences is present.

In some cases, the order of the domains in the biosensor may be critical for appropriate targeting of product(s) and/or reactant to the appropriate cellular compartment(s). For example, the targeting of products or reactants to the peroxisome requires that the peroxisomal targeting domain comprise the last three amino acids of the protein. Determination of those biosensor in which the relative placement of targeting domains within the biosensor is critical can be determined by one of skill in the art through routine experimentation.

Some examples of the basic organization of domains within the protease biosensor are shown in **Figure 30**. One of skill in the art will recognize that any one of a wide variety of protease recognition sites, product target sequences, polypeptide signals, and/or product target sequences can be used in various combinations in the protein biosensor of the present invention, by substituting the appropriate coding sequences into the multi-domain construct. Non-limiting examples of such alternative sequences are shown in **Figure 29A-29C.** Similarly, one of skill in the art will recognize that modifications, substitutions, and deletions can be made to the coding sequences and the amino acid sequence of each individual domain within the biosensor, while retaining the function of the domain. Such various combinations of domains and

modifications

As used herein, the term "coding sequence" or a sequence which encodes a particular polypeptide sequence, refers to a nucleic acid sequence which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide in vitro

or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

As used herein, the term DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell. Not all of these control sequences need always be present in a recombinant vector so long as the DNA sequence of interest is capable of being transcribed and translated appropriately.

As used herein, the term "operatively linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, control sequences operatively linked to a coding sequence are capable of effecting the expression of the coding sequence. The control sequences need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operatively linked" to the coding sequence.

Furthermore, a nucleic acid coding sequence is operatively linked to another nucleic acid coding sequences when the coding region for both nucleic acid molecules are capable of expression in the same reading frame. The nucleic acid sequences need not be contiguous, so long as they are capable of expression in the same reading frame.

[illegible] considered operatively linked

The intervening coding sequences between the various domains of the biosensors can be of any length so long as the function of each domain is retained

Generally, this requires that the two-dimensional and three-dimensional structure of the intervening protein sequence does not preclude the binding or interaction requirements of the domains of the biosensor, such as product or reactant targeting, binding of the protease of interest to the biosensor, fluorescence or luminescence of the detectable polypeptide signal, or binding of fluorescently labeled epitope-specific antibodies.

One case where the distance between domains of the protease biosensor is important is where the goal is to create a fluorescence resonance energy transfer pair. In this case, the FRET signal will only exist if the distance between the donor and acceptor is sufficiently small as to allow energy transfer (Tsien, Heim and Cubbit, WO 97/28261). The average distance between the donor and acceptor moieties should be between 1 nm and 10 nm with a preference of between 1 nm and 6 nm. This is the physical distance between donor and acceptor. The intervening sequence length can vary considerably since the three dimensional structure of the peptide will determine the physical distance between donor and acceptor.

"Recombinant expression vector" includes vectors that operatively link a nucleic acid coding region or gene to any promoter capable of effecting expression of the gene product. The promoter sequence used to drive expression of the protease biosensor may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In a preferred embodiment, the expression vector comprises a plasmid. However, the invention is intended to include any other suitable expression vectors, such as viral vectors.

The phrase "substantially similar " is used herein in reference to the nucleotide sequence of DNA, or the amino acid sequence of protein, having one or more

herein the resulting nucleic acid and/or amino acid sequence is functionally equivalent to the sequences disclosed and claimed herein. Functionally equivalent sequences will function in substantially the same manner to produce substantially the

same protease biosensor as the nucleic acid and amino acid compositions disclosed and claimed herein. For example, functionally equivalent DNAs encode protease biosensors that are the same as those disclosed herein or that have one or more conservative amino acid variations, such as substitutions of non-polar residues for other non-polar residues or charged residues for similarly charged residues, or addition to/deletion from regions of the protease biosensor not critical for functionality. These changes include those recognized by those of skill in the art as substitutions, deletions, and/or additions that do not substantially alter the tertiary structure of the protein.

As used herein, substantially similar sequences of nucleotides or amino acids share at least about 70%-75% identity, more preferably 80-85% identity, and most preferably 90-95% identity. It is recognized, however, that proteins (and DNA or mRNA encoding such proteins) containing less than the above-described level of homology (due to the degeneracy of the genetic code) or that are modified by conservative amino acid substitutions (or substitution of degenerate codons) are contemplated to be within the scope of the present invention.

The term "heterologous" as it relates to nucleic acid sequences such as coding sequences and control sequences, denotes sequences that are not normally associated with a region of a recombinant construct, and/or are not normally associated with a particular cell. Thus, a "heterologous" region of a nucleic acid construct is an identifiable segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, a host cell transformed with a construct which is not normally present in the host cell would be considered heterologous for purposes of this invention.

Manual (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), "Guide to Protein Purification" in

Methods in Enzymology (M.P. Deutshcer, ed., (1990) Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA), *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY), *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, TX).

The biosensors of the present invention are constructed and used to transfect host cells using standard techniques in the molecular biological arts. Any number of such techniques, all of which are within the scope of this invention, can be used to generate protease biosensor-encoding DNA constructs and genetically transfected host cells expressing the biosensors. The non-limiting examples that follow demonstrate one such technique for constructing the biosensors of the invention.

EXAMPLE OF PROTEASE BIOSENSOR CONSTRUCTION AND USE:

In the following examples, caspase-specific biosensors with specific product target sequences have been constructed using sets of 4 primers (2 sense and 2 antisense). These primers have overlap regions at their termini, and are used for PCR via a primer walking technique. (Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) The two sense primers were chosen to start from the 5' polylinker (BspI) of the GFP-containing vector (Clontech, California) to the middle of the designed biosensor sequence. The two antisense primers start from a 3' GFP vector site (Bam HI), and overlap with the sense primers by 12 nucleotides in the middle.

PCR conditions were as follows: 94°C for 30 seconds for denaturation, 55°C for 30 seconds for annealing, and 72°C for 30 seconds for extension for 15 cycles. The primers have restriction endonuclease sites at both ends, facilitating subsequent cloning of the resulting PCR product.

[illegible] resulting fragment was [illegible]

Similarly, the GFP vector (Clontech, San Francisco, CA) was digested at BspE1 and BamH1 sites in the polylinker. Ligation of the GFP vector and the PCR product was performed using standard techniques at 14°C overnight. [illegible] cells were transfected

with the ligation mixtures using standard techniques. Transformed cells were selected on LB-agar with an appropriate antibiotic.

Cells and transfections. For DNA transfection, BHK cells and MCF-7 cells were cultured to 50-70% confluence in 6 well plates containing 3 ml of minimal Eagle's medium (MEM) with 10% fetal calf serum, 1 mM L-glutamine, 50 μg/ml streptomycin, 50 μg/ml penicillin, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate and 10 μg/ml of bovine insulin (for MCF-7 cell only) at 37 °C in a 5% CO_2 incubator for about 36 hours. The cells were washed with serum free MEM media and incubated for 5 hours with 1 ml of transfection mixture containing 1 μg of the appropriate plasmid and 4 μg of lipofectimine (BRL) in the serum free MEM media. Subsequently, the transfection medium was removed and replaced with 3 ml of normal culture media. The transfected cells were maintained in growth medium for at least 16 hours before performing selection of the stable cells based on standard molecular biology methods (Ausubel. et al 1995).

Apoptosis assay. For apoptosis assays, the cells (BHK, MCF-7) stably transfected with the appropriate protease biosensor expression vector were plated on tissue culture treated 96-well plates at 50-60% confluence and cultured overnight at 37°C, 5% CO_2. Varying concentrations of cis-platin, staurosporine, or paclitaxel in normal culture media were freshly prepared from stock and added to cell culture dishes to replace the old culture media. The cells were then observed with the cell screening system of the present invention at the indicated time points either as live cell experiments or as fixed end-point experiments.

1. **Construction of 3-domain protease biosensors**

a. **Caspase-3 biosensor with an annexin II reactant targeting domain**

[illegible] biosensor [illegible] Figure 5A, and its sequence is shown in SEQ ID NO:1 and 2.

Primers for Caspase 3, Product target sequence = none (CP3GFP-CYTO):

1) TCA TCA TCC GGA GCT GGA GCC GGA GCT GGC CGA TCG GCT GTT AAA TCT GAA GGA AAG AGA AAG TGT GAC GAA GTT GAT GGA ATT GAT GAA GTA GCA (**SEQ ID NO:153**)

2) GAA GAA GGA TCC GGC ACT TGG GGG TGT AGA ATG AAC ACC CTC CAA GCT GAG CTT GCA CAG GAT TTC GTG GAC AGT AGA CAT AGT ACT TGC TAC TTC ATC (**SEQ ID NO:154**)

3) TCA TCA TCC GGA GCT GGA (**SEQ ID NO:155**)

4) GAA GAA GGA TCC GGC ACT (**SEQ ID NO:156**)

This biosensor is restricted to the cytoplasm by the reactant target sequence. The reactant target sequence is the annexin II cytoskeletal binding domain (MSTVHEILCKLSLEGVHSTPPSA) **(SEQ ID NO:124)** **(Figure 29C)** (Eberhard et al. 1997. *Mol. Biol. Cell* 8:293a). The enzyme recognition site corresponds to two copies of the amino acid sequence DEVD **(SEQ ID NO:60)** **(Figure 29B)**, which serves as the recognition site of caspase-3. Other examples with different numbers of protease recognition sites and/or additional amino acids from a naturally occurring protease recognition site are shown below. The signal domain is EGFP **(SEQ ID NO:46)** **(Figure 29A)** (Clontech, California). The parent biosensor (the reactant) is restricted to the cytoplasm by binding of the annexin II domain to the cytoskeleton, and is therefore excluded from the nucleus. Upon cleavage of the protease recognition site by caspase 3, the signal domain (EGFP) is released from the reactant targeting domain (annexin II) and is distributed throughout the [illegible]

[illegible]g 33

The biosensor response is measured by quantitating the effective cytoplasm-to-nuclear translocation of the signal (see above). Measurement of the response is by one

of several modes, including integrated or average nuclear region intensity, the ratio or difference of the integrated or average cytoplasm intensity to integrated or average nuclear intensity. The nucleus is defined using a DNA-specific dye, such as Hoechst 33342.

This biosensor provides a measure of the proteolytic activity around the annexin II cytoskeleton binding sites within the cell. Given the dispersed nature of the cytoskeleton and the effectively diffuse state of cytosolic enzymes, this provides an effective measure of the cytoplasm in general.

Results & Discussion:

Fig 33 illustrates images before and after stimulation of apoptosis by cis-platin in BHK cells, transfected with the caspase 3 biosensor. The images clearly illustrate accumulation of fluorescence in the nucleus. Generation of the spatial change in fluorescence is non-reversible and thus the timing of the assay is flexible. Controls for this biosensor include using a version in which the caspase-3-specific site has been omitted. In addition, disruption of the cytoskeleton with subsequent cell rounding did not produce the change in fluorescence distribution. Our experiments demonstrate the correlation of nuclear condensation with activation of caspase activity. We have also tested this biosensor in MCF-7 cells. A recent report measured a peak response in caspase-3 activity 6 h after stimulation of MCF-7 cells with etoposide accompanied by cleavage of PARP (Benjamin et al. 1998.*Mol Pharmacol.* 53:446-50). However, another recent report found that MCF-7 cells do not possess caspase-3 activity and, in fact, the caspase–3 gene is functionally deleted (Janicke et al. 1998. *J Biol Chem.* 273:9357-60). Caspase-3 activity was not detected with the caspase biosensor in MCF-7 cells after a 15 h treatment with 100 μM etoposide.

Janicke et al., (1998) also indicated that many of the conventional substrates of caspase-3 were cleaved in MCF-7 cells upon treatment with staurosporine. Our

activation by staurosporine was approximately one-half that demonstrated with cis-platin in BHK cells. This also implies that the current biosensor, although designed to be caspase-3-specific, is indeed specific for a class of caspases rather than uniquely

specific for caspase-3. The most likely candidate is caspase-7 (Janicke et al., 1998). These experiments also demonstrated that the biosensor can be used in multiparameter experiments, with the correlation of decreases in mitochondrial membrane potential, nuclear condensation, and caspase activation.

We have specifically tested the effects of paclitaxel on caspase activation using the biosensor. Caspase activity in BHK and MCF-7 cells was stimulated by paclitaxel. It also appears that caspase activation occurred after nuclear morphology changes. One caveat is that, based on the above discussions, the caspase activity reported by the biosensor in this assay is likely to be due to the combination of caspase-3 and, at least, caspase-7 activity.

Consistent with the above results using staurosporine stimulation on MCF-7 cells, paclitaxel also stimulated the activation of caspase activity. The magnitude was similar to that of staurosporine. This experiment used a much narrower range of paclitaxel than previous experiments where nuclear condensation appears to dominate the response.

b. Caspase biosensor with the microtubule associated protein 4 (MAP4) projection domain (CP8GFPNLS-SIZEPROJ)

Another approach for restricting the reactant to the cytoplasm is to make the biosensor too large to penetrate the nuclear pores Cleavage of such a biosensor liberates a product capable of diffusing into the nucleus.

The additional size required for this biosensor is provided by using the projection domain of MAP4 **(SEQ ID NO:142) (Figure 29C)** (CP8GFPNLS-SIZEPROJ). The projection domain of MAP4 does not interact with microtubules on its own, and, when expressed, is diffusely distributed throughout the cytoplasm, but is excluded from the nucleus due to its size (~120 kD). Thus, this biosensor is distinct from the one using the full length MAP4 sequence. (see below) One of skill in the art will recognize that many other [illegible]

[illegible] copies of any other protein that lacks an active NLS and exceeds the maximum size for diffusion into the nucleus (approximately 60 kD; Alberts, B., Bray, D., Raff, M., Roberts, K., Watson, J.D. (Eds.) Molecular Biology of the Cell, third edition, New

York: Garland publishing, 1994. pp 561-563). The complete sequence of the resulting biosensor is shown in **Figure 34. (SEQ ID NO: 3-4)** A similar biosensor with a different protease recognition domain is shown in **Figure 35 (SEQ ID NO:5-6)**

c. Caspase biosensor with a nuclear export signal

Another approach for restricting the reactant to the cytoplasm is to actively restrict the reactant from the nucleus by using a nuclear export signal. Cleavage of such a biosensor liberates a product capable of diffusing into the nucleus.

The Bacillus anthracis bacterium expresses a zinc metalloprotease protein complex called anthrax protease. Human mitogen activated protein kinase kinase 1 (MEK 1) (Seger et al., J. Biol. Chem. 267:25628-25631, 1992) possesses an anthrax protease recognition site (amino acids 1-13) **(SEQ ID NO:102) (Figure 29B)** that is cleaved after amino acid 8, as well as a nuclear export signal at amino acids **32-44 (SEQ ID NO:140) (Figure 29C).** Human MEK 2 (Zheng and Guan, J. Biol. Chem. 268:11435-11439, 1993) possesses an anthrax protease recognition site comprising amino acid residues 1-16 **(SEQ ID NO:104) (Figure 29B)** and a nuclear export signal at amino acids 36-48. **(SEQ ID NO:148) (Figure 29C).**

The anthrax protease biosensor comprises Fret25 **(SEQ ID NO:48) (Figure 29A)** as the signal, the anthrax protease recognition site, and the nuclear export signal from MEK 1 or MEK2. **(SEQ ID NOS: 7-8 (MEK1; Fig. 36); 9-10 (MEK2) Fig. 37)** The intact biosensor will be retained in the cytoplasm by virture of this nuclear export signal (eg., the reactant target site). Upon cleavage of the fusion protein by anthrax protease, the NES will be separated from the GFP allowing the GFP to diffuse into the nucleus.

2. Construction of 4- and 5-domain biosensors

For all of the examples presented above for 3-domain protease biosensors, a

... thus cause the signal sequence to segregate from the reactant target domain after proteolytic cleavage. Addition of a second detectable signal domain, including but not limited to those in **Figure 29A**, operatively linked with the reactant target domain is

also useful in allowing measurement of the reaction by multiple means. Specific examples of such biosensors are presented below.

a. 4 domain biosensors

1. Caspase biosensors with nuclear localization sequences (pcas3nlsGFP; CP3GFPNLS-CYTO):

The design of the biosensor is outlined in **Figure 38,** and its sequence is shown in **SEQ ID NO:11-12 (Figure 39).** PCR and cloning procedures were performed as described above, except that the following oligonucleotides were used:

Primers for Caspase 3, Product target sequence = NLS (CP3GFPNLS-CYTO) :

1) TCA TCA TCC GGA AGA AGG AAA CGA CAA AAG CGA TCG GCT GTT AAA TCT GAA GGA AAG AGA AAG TGT GAC GAA GTT GAT GGA ATT GAT GAA GTA GCA **(SEQ ID NO:157)**

2) GAA GAA GGA TCC GGC ACT TGG GGG TGT AGA ATG AAC ACC CTC CAA GCT GAG CTT GCA CAG GAT TTC GTG GAC AGT AGA CAT AGT ACT TGC TAC TTC ATC **(SEQ ID NO:154)**

3) TCA TCA TCC GGA AGA AGG **(SEQ ID NO:158)**

4) GAA GAA GGA TCC GGC ACT **(SEQ ID NO:156)**

This biosensor is similar to that shown in **SEQ ID NO:2** except upon recognition and cleavage of the protease recognition site, the product is released and the signal accumulates specifically in the nucleus due to the presence of a nuclear localization sequence, RRKRQK **(SEQ ID NO:128) (Figure 29C)**(Briggs et al., J. Biol. Chem. 273:22715, 1998) [illegible]

[illegible] in the cytoplasm where the product of the enzymatic reaction is restricted to the nuclear compartment. The response is measured by quantitating the effective cytoplasm-to-nuclear translocation of the signal, as described above.

With the presence of both product and reactant targeting sequences in the parent biosensor, the reactant target sequence should be dominant prior to activation (e.g., protease cleavage) of the biosensor. One way to accomplish this is by masking the product targeting sequence in the parent biosensor until after protease cleavage. In one such example, the product target sequence is functional only when relatively near the end of a polypeptide chain (ie: after protease cleavage). Alternatively, the biosensor may be designed so that its tertiary structure masks the function of the target sequence until after protease cleavage. Both of these approaches include comparing targeting sequences with different relative strengths for targeting. Using the example of the nuclear localization sequence (NLS) and annexin II sequences, different strengths of NLS have been tried with clone selection based on cytoplasmic restriction of the parent biosensor. Upon activation, the product targeting sequence will naturally dominate the localization of its associated detectable sequence domain because it is then separated from the reactant targeting sequence.

An added benefit of using this biosensor is that the product is targeted, and thus concentrated, into a smaller region of the cell. Thus, smaller amounts of product are detectable due to the increased concentration of the product. This concentration effect is relatively insensitive to the cellular concentration of the reactant. The signal-to-noise ratio (SNR) of such a measurement is improved over the more dispersed distribution of biosensor #1.

Similar biosensors that incorporate either the caspase 6 (**SEQ ID NO:66**) (**Figure 29B**) or the caspase 8 protease recognition sequence (**SEQ ID NO:74**) (**Figure 29B**) can be made using the methods described above, but using the following primer sets:

Primers for Caspase 6, Product target sequence = NLS (CP6GFPNLS-CYTO)

1) TCA TCA TCC GGA AGA AGG AAA CGA CAA AAG CGA TCG

[illegible] CAA GCT GAG CTT GCA CAG GAT TTC GTG GAC

AGT AGA CAT AGT ACT GTT GTC AAT TTC (**SEQ ID NO:160**)

3) TCA TCA TCC GGA AGA AGG (**SEQ ID NO:158**)

4) GAA GAA GGA TCC GGC ACT (**SEQ ID NO:156**)

Primers for Caspase 8, Product target sequence = NLS (CP8GFPNLS-CYTO)

1) TCA TCA TCC GGA AGA AGG AAA CGA CAA AAG CGA TCG TAT CAA AAA GGA ATA CCA GTT GAA ACA GAC AGC GAA GAG CAA CCT TAT (**SEQ ID NO:161**)

2) GAA GAA GGA TCC GGC ACT TGG GGG TGT AGA ATG AAC ACC CTC CAA GCT GAG CTT GCA CAG GAT TTC GTG GAC AGT AGA CAT AGT ACT ATA AGG TTG CTC (**SEQ ID NO:162**)

3) TCA TCA TCC GGA AGA AGG (**SEQ ID NO:158**)

4) GAA GAA GGA TCC GGC ACT (**SEQ ID NO:156**)

The sequence of the resulting biosensors is shown in **Figures 40 (Caspase 6) (SEQ ID NO:13-14)** and **41 (Caspase 8) (SEQ ID NO: 15-16).** Furthermore, multiple copies of the protease recognition sites can be inserted into the biosensor, yielding the biosensors shown in **Figures 42 (Caspase 3) (SEQ ID NO: 17-18)** and **43 (Caspase 8) (SEQ ID NO:19-20).**

2. Caspase 3 biosensor with a second signal domain

An alternative embodiment employs a second signal domain operatively linked to the reactant target domain. In this example, full length MAP4 serves as the reactant target sequence. Upon recognition and cleavage, one product of the reaction, containing the reactant target sequence, remains bound to microtubules in the cytoplasm with its own unique signal, while the other product, containing the product target sequence, diffuses into the nucleus. This biosensor provides a means to measure two activities at once: caspase 3 activity using a translocation of GFP into the nucleus and microtubule [illegible]

The basic premise for this biosensor is that the reactant is tethered to the microtubule cytoskeleton by virtue of the reactant target sequence comprising the full length microtubule associated protein MAP4 (**SEQ ID NO:152**) (**Figure 29C**). In this

case, a DEVD (**SEQ ID NO:60**) (**Figure 29B**) recognition motif is located between the EYFP signal (**SEQ ID NO:44**) (**Figure 29A**) operatively linked to the reactant target sequence, as well as the EBFP signal (**SEQ ID NO:48**) (**Figure 29A**) operatively linked to the C-terminus of MAP4. The resulting biosensor is shown in **Figure 44. (SEQ ID NO:21-22)**

This biosensor can also include a product targeting domain, such as an NLS, operatively linked to the signal domain.

With this biosensor, caspase-3 cleavage still releases the N-terminal GFP, which undergoes translocation to the nucleus (directed there by the NLS). Also, the MAP4 fragment, which is still intact following proteolysis by caspase-3, continues to report on the integrity of the microtubule cytoskeleton during the process of apoptosis via the second GFP molecule fused to the C-terminus of the biosensor. Therefore, this single chimeric protein allows simultaneous analysis of caspase-3 activity and the polymerization state of the microtubule cytoskeleton during apoptosis induced by a variety of agents. This biosensor is also useful for analysis of potential drug candidates that specifically target the microtubule cytoskeleton, since one can determine whether a particular drug induced apoptosis in addition to affecting microtubules.

This biosensor potentially combines a unique signal for the reactant, fluorescence resonance energy transfer (FRET) from signal 2 to signal 1, and a unique signal localization for the product, nuclear accumulation of signal 1. The amount of product generated will also be indicated by the magnitude of the loss in FRET, but this will be a smaller SNR than the combination of FRET detection of reactant and spatial localization of the product.

FRET can occur when the emission spectrum of a donor overlaps significantly the absorption spectrum of an acceptor molecule. (dos Remedios, C.G., and P.D. Moens. 1995. Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. *J Struct Biol.* 115:175-85. [illegible]

[illegible] concentration changes at the secretory vesicle surface with a recombinant targeted cameleon. *Curr Biol.* 9:915-918.) The average physical distance between the donor and acceptor molecules should be between 1 nm and 10 nm with a preference of between 1

nm and 6 nm. The intervening sequence length can vary considerably since the three dimensional structure of the peptide will determine the physical distance between donor and acceptor. This FRET signal can be measured as (1) the amount of quenching of the donor in the presence of the acceptor, (2) the amount of acceptor emission when exciting the donor, and/or (3) the ratio between the donor and acceptor emission. Alternatively, fluorescent lifetimes of donor and acceptor could be measured.

This case adds value to the above FRET biosensor by nature of the existence of the reactant targeting sequence. This sequence allows the placement of the biosensor into specific compartments of the cell for a more direct readout of activity in those compartments such as the inner surface of the plasma membrane.

The cytoplasmic second signal represents both original reactant plus one part of the product. The nuclear first signal represents another product of the reaction. Thus the enzymatic reaction has the added flexibility in that it can be represented as (1) nuclear intensity; (2) the nucleus /cytoplasm ratio; (3) the nucleus /cytoplasm FRET ratio; (4) cytoplasmic /cytoplasmic FRET ratio.

The present FRET biosensor design differs from previous FRET-based biosensors (see WO 97/28261; WO9837226) in that it signal measurement is based on spatial position rather than intensity. The products of the reaction are segregated from the reactants. It is this change in spatial position that is measured. The FRET-based biosensor is based on the separation, but not to another compartment, of a donor and acceptor pair. The intensity change is due to the physical separation of the donor and acceptor upon proteolytic cleavage. The disadvantages of FRET-based biosensors are (1) the SNR is rather low and difficult to measure, (2) the signal is not fixable. It must be recorded using living cells. Chemical fixation, for example with formaldehyde, cannot preserve both the parent and resultant signal; (3) the range of wavelengths are limiting and cover a larger range of the spectrum due to the presence of two fluorophores or a fluorophore and chromophore; (4) the construction has greater limitations in that the donor and

Benefits of the positional biosensor includes: (1) ability to concentrate the signal in order to achieve a higher SNR. (2) ability to be used with either living or fixed cells; (3) only a single fluorescent signal is needed; (4) the arrangement of the domains

of the biosensor is more flexible. The only limiting factor in the application of the positional biosensor is the need to define the spatial position of the signal which requires an imaging method with sufficient spatial resolution to resolve the difference between the reactant compartment and the product compartment.

One of skill in the art will recognize that this approach can be adapted to report any desired combination of activities by simply making the appropriate substitutions for the protease recognition sequence and the reactant target sequence, including but not limited to those sequences shown in **Figure 29A-C.**

3. Caspase 8 biosensor with a nucleolar localization domain (CP8GFPNUC-CYTO)

This approach (diagrammed in **Figure 45**) utilizes a biosensor for the detection of caspase-8 activity. In this biosensor, a nucleolar localization signal (RKRIRTYLKSCRRMKRSGFEMSRPIPSHLT) **(SEQ ID NO:130)** (**Figure 29C**) (Ueki et al., Biochem. Biophys. Res. Comm. 252:97-100, 1998) was used as the product target sequence, and made by PCR using the primers described below. The PCR product was digested with BspE1 and Pvu1 and gel purified. The vector and the PCR product were ligated as described above.

Primers for Caspase 8, Nucleolar localization signal (CP8GFPNUC-CYTO):

1) TCA TCA TCC GGA AGA AAA CGT ATA CGT ACT TAC CTC AAG TCC TGC AGG CGG ATG AAA AGA (SEQ ID NO:163)
2) GAA GAA CGA TCG AGT AAG GTG GGA AGG AAT AGG TCG AGA CAT CTC AAA ACC ACT TCT TTT CAT (SEQ ID NO:164)
3) TCA TCA TCC GGA AGA AAA (SEQ ID NO:165)
4) GAA GAA CGA TCG AGT AAG (SEQ ID NO:166)

23 24) This biosensor includes the protease recognition site for caspase-8 (**SEQ ID NO:74**) (**Figure 29B**). A similar biosensor utilizes the protease recognition site for caspase-3 (**Figure 47**; SEQ ID NO:25-26)

These biosensors could be used with other biosensors that possess the same product signal color that are targeted to separate compartments, such as CP3GFPNLS-CYTO. The products of each biosensor reaction can be uniquely measured due to separation of the products based on the product targeting sequences. Both products from CP8GFPNUC-CYTO and CP3GFPNLS-CYTO are separable due to the different spatial positions, nucleus vs. nucleolus, even though the colors of the products are exactly the same. Assessing the non-nucleolar, nuclear region in order to avoid the spatial overlap of the two signals would perform the measurement of CP3GFPNLS in the presence of CP8GFPNUC. The loss of the nucleolar region from the nuclear signal is insignificant and does not significantly affect the SNR. The principle of assessing multiple parameters using the same product color significantly expands the number of parameters that can be assessed simultaneously in living cells. This concept can be extended to other non-overlapping product target compartments.

Measurement of translocation to the nucleolar compartment is performed by (1) defining a mask corresponding to the nucleolus based on a nucleolus-specific marker, including but not limited to an antibody to nucleolin (Lischwe et al., 1981. *Exp. Cell Res.* 136:101-109); (2) defining a mask for the reactant target compartment, and (3) determining the relative distribution of the signal between these two compartments. This relative distribution could be represented by the difference in the two intensities or, preferably, the ratio of the intensities between compartments.

The combination of multiple positional biosensors can be complicated if the reactant compartments are overlapping. Although each signal could be measured by simply determining the amount of signal in each product target compartment, higher SNR will be possible if each reactant is uniquely identified and quantitated. This higher SNR can be maximized by adding a second signal domain of contrasting fluorescent property. This second signal may be produced by a signal domain operatively linked to the product targeting sequence, or by FRET (see above), or by a reactant targeting sequence [illegible]

[illegible] Alternatively, for large compartments, such as the cytoplasm, it is possible to place different, same colored biosensors in different parts of the same compartment.

4. Protease biosensors with multiple copies of a second signal domain serving as a reactant target domain

In another example, (CP8YFPNLS-SIZECFPn) increasing the size of the reactant is accomplished by using multiple inserts of a second signal sequence, for example, ECFP (**SEQ ID NO:50**) (**Figure 29A**) (Tsien, R.Y. 1998. Annu Rev Biochem. 67:509-44). Thus, the multiple copies of the second signal sequence serve as the reactant target domain by excluding the ability of the biosensor to diffuse into the nucleus. This type of biosensor provides the added benefit of additional signal being available per biosensor molecule. Aggregation of multiple fluorescent probes also can result in unique signals being manifested, such as FRET, self quenching, eximer formation, etc. This could provide a unique signal to the reactants.

5. Tetanus/botulinum biosensor with trans-membrane targeting domain

In an alternative embodiment, a trans-membrane targeting sequence is used to tether the reactant to cytoplasmic vesicles, and an alternative protease recognition site is used. The tetanus/botulinum biosensor (**FIG NOS. 48-49**) (**SEQ ID NOS:27-28** (cellubrevin); **29-30** (synaptobrevin) consists of an NLS (**SEQ ID NO:128**) (**Figure 29C**), Fret25 signal domain (**SEQ ID NO:52**) (**Figure 29A**), a tetanus or botulinum zinc metalloprotease recognition site from cellubrevin (**SEQ ID NO:106**) (**Figure 29B**) (McMahon et al., Nature 364:346-349, 1993; Martin et al., J. Cell Biol., in press) or synaptobrevin (**SEQ ID NO:108**) (**Figure 29B**) (GenBank Accession #U64520), and a trans-membrane sequence from cellubrevin (**SEQ ID NO:146**) (**Figure 29C**) or synaptobrevin (**SEQ ID NO:144**) (**Figure 29C**) at the 3'-end which tethers the biosensor to cellular vesicles. The N-terminus of each protein is oriented towards the cytoplasm. In the intact biosensor, GFP is tethered to the vesicles. Upon cleavage by the tetanus or botulinum zinc metalloprotease, GFP will no longer be associated with

domain biosensors

1. Caspase 3 biosensor with a nuclear localization domain and a second signal domain operatively linked to an annexin II domain

The design of this biosensor is outlined in **Figure 50,** and the sequence is shown in **Figure 52 (SEQ ID NO:33-34).** This biosensor differs from **SEQ ID NO 11-12** by including a second detectable signal, ECFP **(SEQ ID NO:50) (Figure 29A)** (signal 2) operatively linked to the reactant target sequence.

2. Caspase 3 biosensor with a nuclear localization sequence and a second signal domain operatively linked to a MAP4 projection domain (CP3YFPNLS-CFPCYTO)

In this biosensor **(Figure 51) (SEQ ID NO:31-32)**, an NLS product targeting domain **(SEQ ID NO:128) (Figure 29C)** is present upstream of an EYFP signal domain **(SEQ ID NO:44) (Figure 29A)**. A DEVD protease recognition domain **(SEQ ID NO:60) (Figure 29B)** is between after the EYFP signal domain and before the MAP4 projection domain **(SEQ ID NO:142) (Figure 29C)**.

While a preferred form of the invention has been shown in the drawings and described, since variations in the preferred form will be apparent to those skilled in the art, the invention should not be construed as limited to the specific form shown and described, but instead is as set forth in the claims.

CLAIMS

We claim:

1. A recombinant nucleic acid encoding a protease biosensor, comprising:
   a. a first nucleic acid sequence that encodes at least one detectable polypeptide signal;
   b. a second nucleic acid sequence that encodes at least one protease recognition site, wherein the second nucleic acid sequence is operatively linked to the first nucleic acid sequence that encodes the at least one detectable polypeptide signal; and
   c. a third nucleic acid sequence that encodes at least one reactant target sequence, wherein the third nucleic acid sequence is operatively linked to the second nucleic acid sequence that encodes the at least one detectable polypeptide signal.

2. The recombinant nucleic acid biosensor of claim 1 further comprising a fourth nucleic acid sequence that encodes at least one product target sequence, wherein the fourth nucleic acid sequence is operatively linked to the first nucleic acid sequence that encodes the at least one detectable polypeptide signal.

3. The recombinant nucleic acid biosensor of claim 1 **or 2** further comprising a fifth nucleic acid sequence that encodes at least one detectable polypeptide signal, wherein the fifth nucleic acid sequence is operatively linked to the third nucleic acid sequence that encodes the reactant target sequence.

4. The recombinant nucleic acid of claim 1-3 wherein the detectable polypeptide signal is selected from the group consisting of fluorescent proteins, luminescent proteins, sequence epitopes, and co-factor requiring fluorescent or luminescent proteins.

5. The recombinant nucleic acid biosensor of claim 1-3 wherein the first nucleic acid encoding a detectable polypeptide sequence comprises a sequence selected from the group consisting of **SEQ ID NOS: 35, 37, 39, 41, 43, 45, 47, 49, and 51.**

6. The recombinant nucleic acid biosensor of claim 1 wherein the second nucleic acid encoding a protease recognition site comprises a sequence selected from the group consisting of **SEQ ID NOS: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, and 121.**

[illegible] group consisting of **SEQ ID NOS:** 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, and 151.

8. A recombinant nucleic acid biosensor encoding a protease biosensor comprising a sequence substantially similar to sequences selected from the group consisting of **SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.**

9. A recombinant expression vector comprising DNA control sequences operatively linked to the recombinant nucleic acid of any one of claim 1-8.

10. A recombinant protease biosensor comprising
    a. a first domain comprising at least one detectable polypeptide signal;
    b. a second domain comprising at least protease recognition site; and
    c. a third domain comprising at least one reactant target sequence;
wherein the first domain and the third domain are separated by the second domain.

11. The recombinant protease biosensor of claim 10 further comprising a fourth domain comprising at least one product target sequence, wherein the fourth domain and the third domain are separated by the second domain.

12. The recombinant protease biosensor of claim 10 or 11 further comprising a fifth domain comprising at least one detectable polypeptide signal, wherein the fifth domain and the first domain are separated by the second domain.

13. The protease biosensor of claim 10-12 wherein the detectable polypeptide signal is selected from the group consisting of fluorescent proteins, luminescent proteins, sequence epitopes, and co-factor requiring fluorescent or luminescent proteins.

14. The protease biosensor of claim 10-12 wherein the detectable polypeptide signal domain comprises a sequence selected from the group consisting of **SEQ ID NOS:36, 38, 40, 42, 44, 46, 48, 50, and 52.**

15. The protease biosensor of claim 10 wherein the second domain comprising a protease recognition site comprises a sequence selected from the group consisting of **SEQ ID NOS:54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, and 122.**

16. The protease biosensor of claim 10 wherein the reactant target sequence domain comprise a sequence selected from the group consisting of **SEQ ID NOS:124.** 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, [illegible]

[illegible] sequences selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, **and 34.**

18. A genetically engineered host cell that has been transfected with the recombinant expression vector of claim 9.

19. An method for identifying compounds that modify protease activity in a cell, comprising
    a) providing a host cell that possesses the recombinant protease biosensor of any one of claim 10-17;
    b) contacting the host cell with a test compound;
    c) determining the protease biosensor distribution in the host cell, wherein changes in the distribution of the protease biosensor are correlated with modification of protease activity by the test compound.

20. A kit for identifying compounds that modify protease activity in a host cell, comprising;
    a) the recombinant nucleic acid encoding a protease biosensor of any one of claims 1-8; and
    b) instructions for use of the recombinant nucleic acid to identify compounds that modify protease activity in a host cell.

21. A kit for identifying compounds that modify protease activity in a host cell, comprising;
    a) the recombinant protease biosensor of any one of claims 10-17; and
    b) instructions for use of the recombinant protease biosensor to identify compounds that modify protease activity in a host cell.

22. A kit for identifying compounds that modify protease activity in a host cell, comprising;
    a) the recombinant expression vector of claim 9; and
    b) instructions for use of the recombinant expression vector to identify compounds that modify protease activity in a host cell.

23. A kit for identifying compounds that modify protease activity in a host cell, comprising;
    a) the genetically engineered host cell of claim 18; and
    b) instructions for use of the genetically engineered host cell to identify compounds that modify protease activity in the host cell.

Figure 1

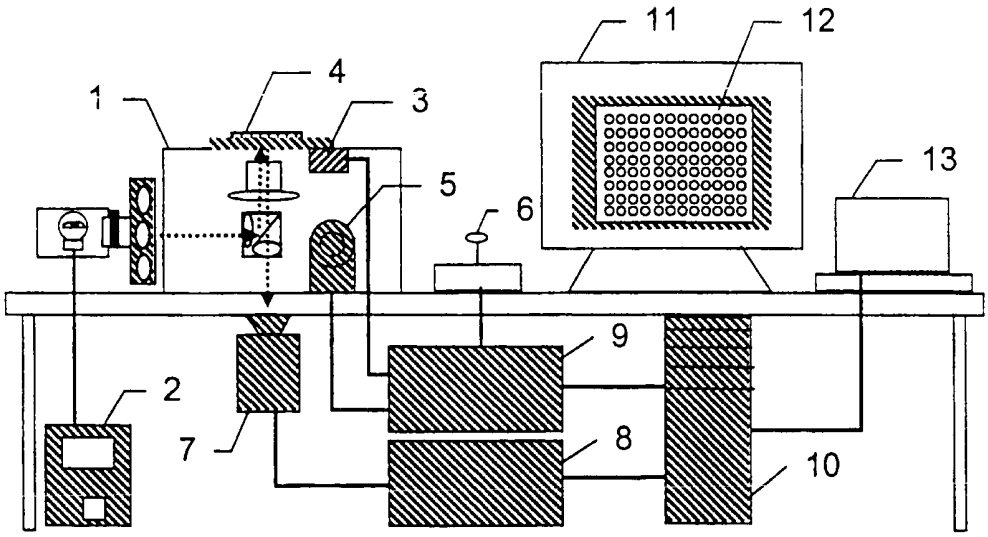
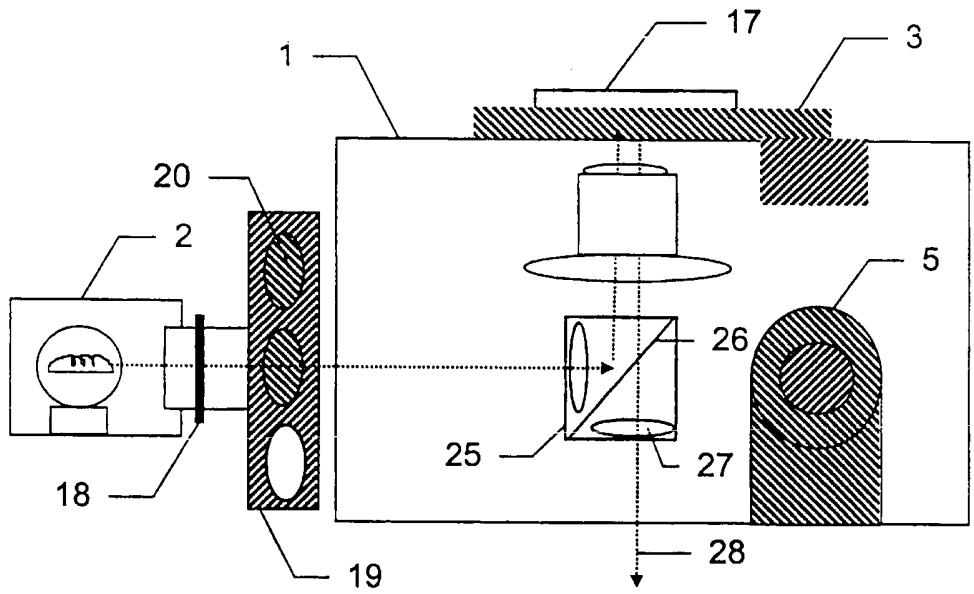


Figure 2



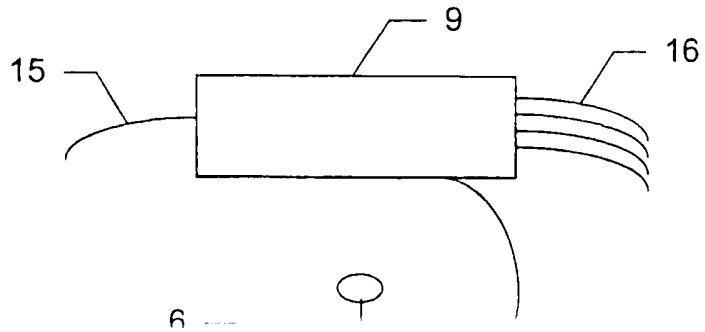


Figure 3

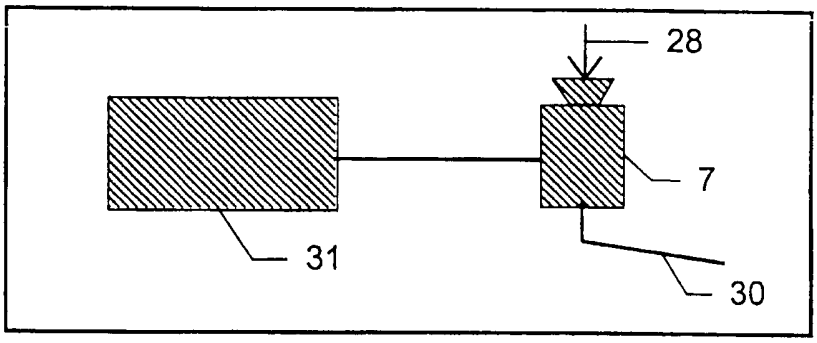


Figure 4

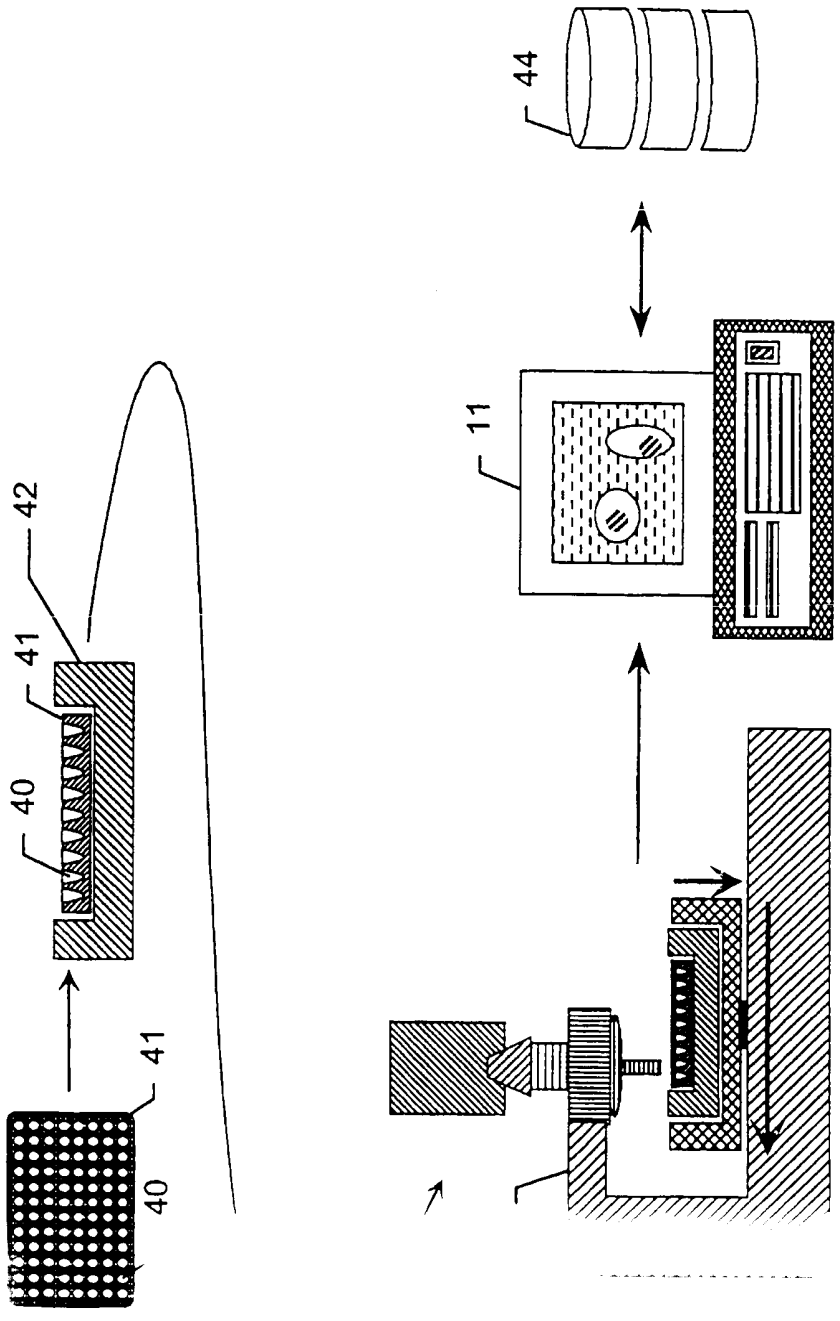


Figure 5

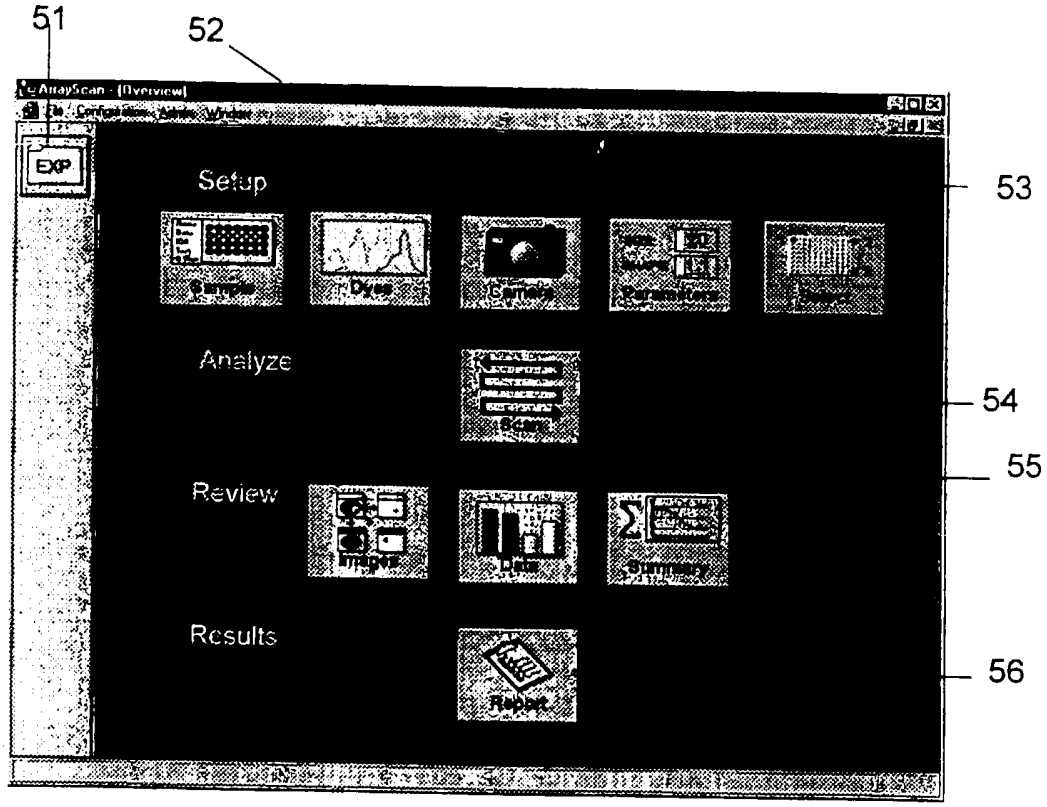


Figure 6

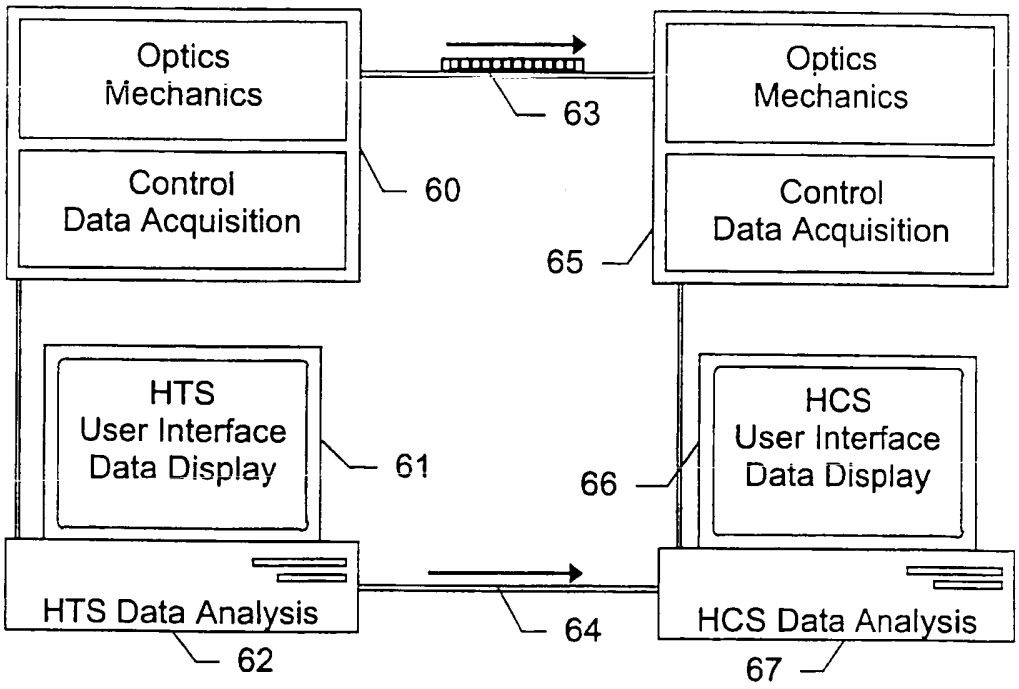

Optics
Mechanics
Control
Data Acquisition
60
63
Optics
Mechanics
Control
Data Acquisition
65
HTS
User Interface
Data Display
61
HCS
User Interface
Data Display
66
HTS Data Analysis
62
64
HCS Data Analysis
67

Figure 7

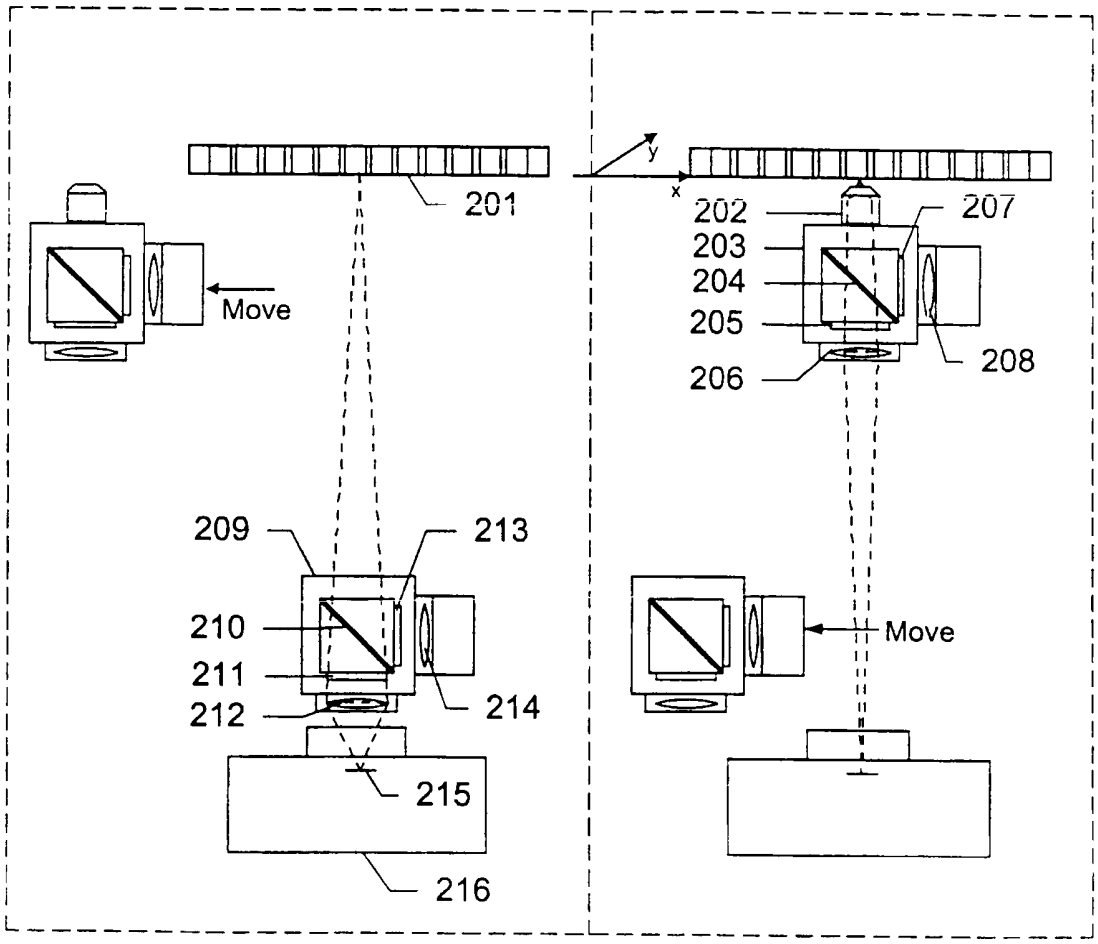

201
Move
209
213
210
211
212
214
215
216
y
x
202
203
204
205
206
207
208
Move

Figure 8

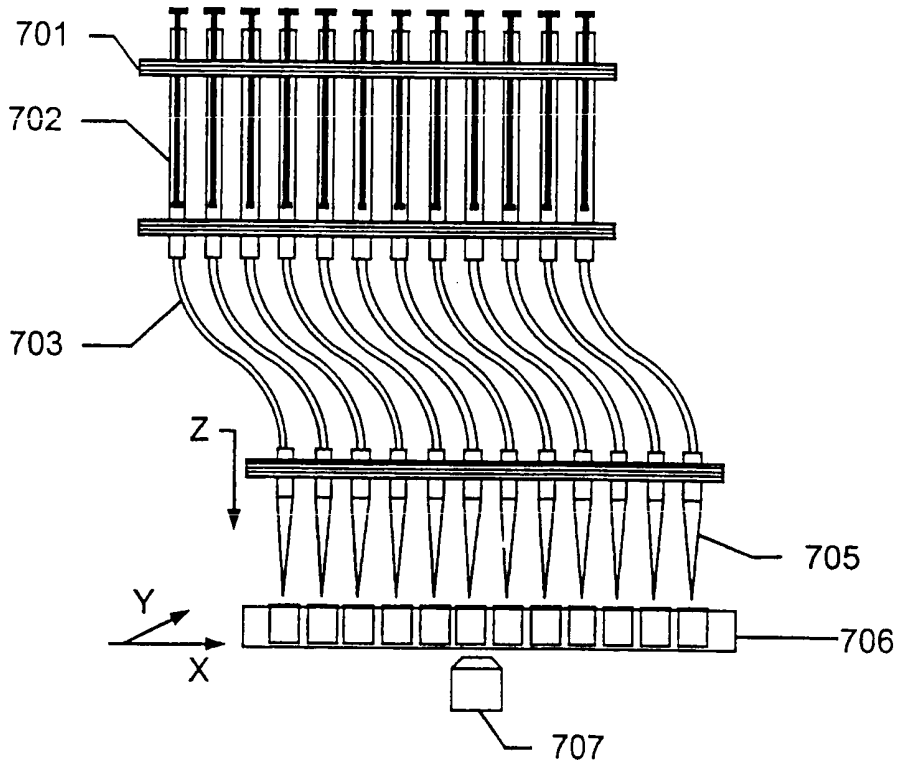


Figure 9

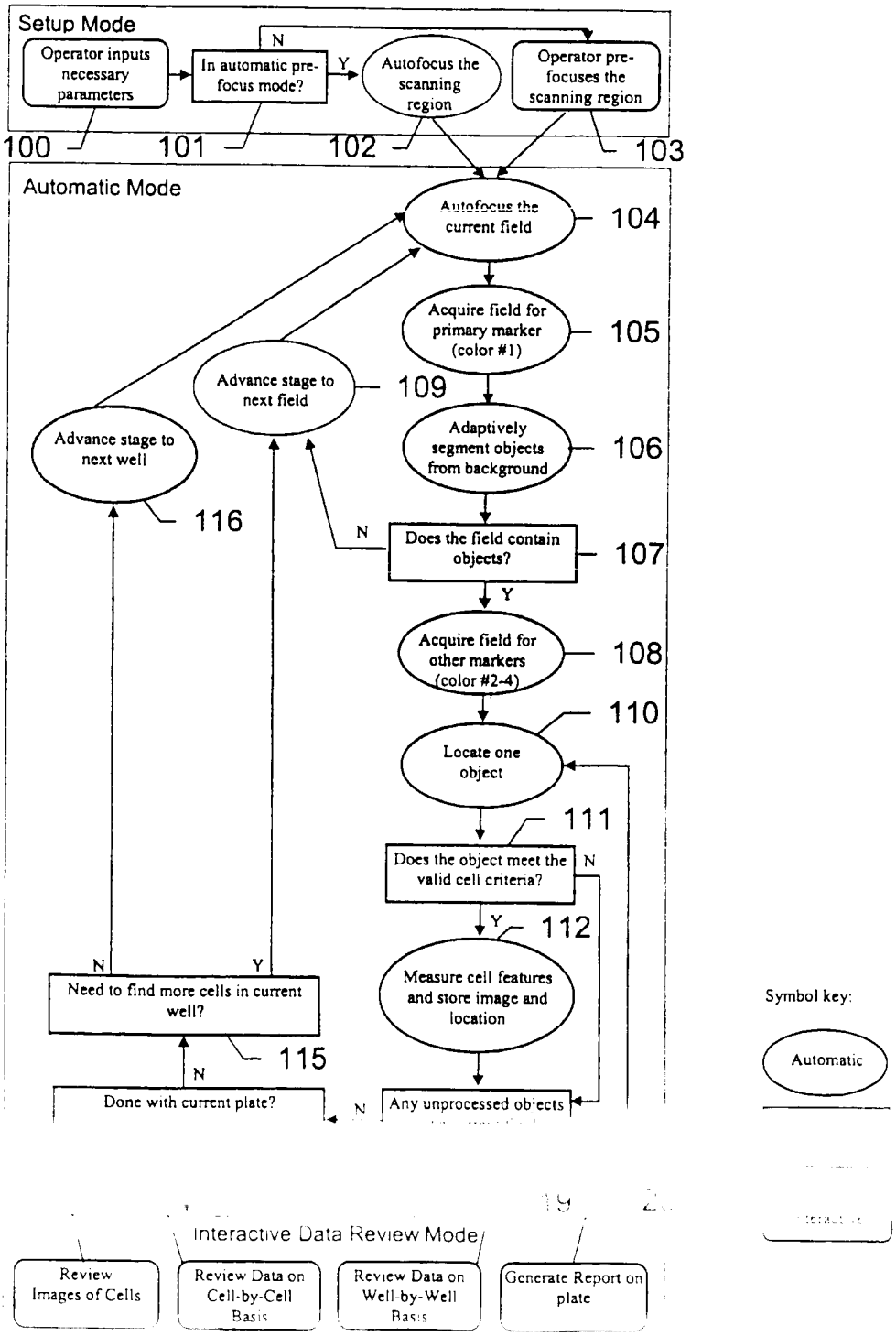

Setup Mode
Operator inputs necessary parameters
In automatic pre-focus mode?
N
Y
Autofocus the scanning region
Operator pre-focuses the scanning region
100
101
102
103
Automatic Mode
Autofocus the current field
104
Acquire field for primary marker (color #1)
105
Advance stage to next field
109
Adaptively segment objects from background
106
Advance stage to next well
116
N
Does the field contain objects?
107
Y
Acquire field for other markers (color #2-4)
108
110
Locate one object
111
Does the object meet the valid cell criteria?
N
Y
112
Measure cell features and store image and location
N
Y
Need to find more cells in current well?
115
N
Done with current plate?
N
Any unprocessed objects
Symbol key:
Automatic
Interactive Data Review Mode
Review Images of Cells
Review Data on Cell-by-Cell Basis
Review Data on Well-by-Well Basis
Generate Report on plate

Figure 10

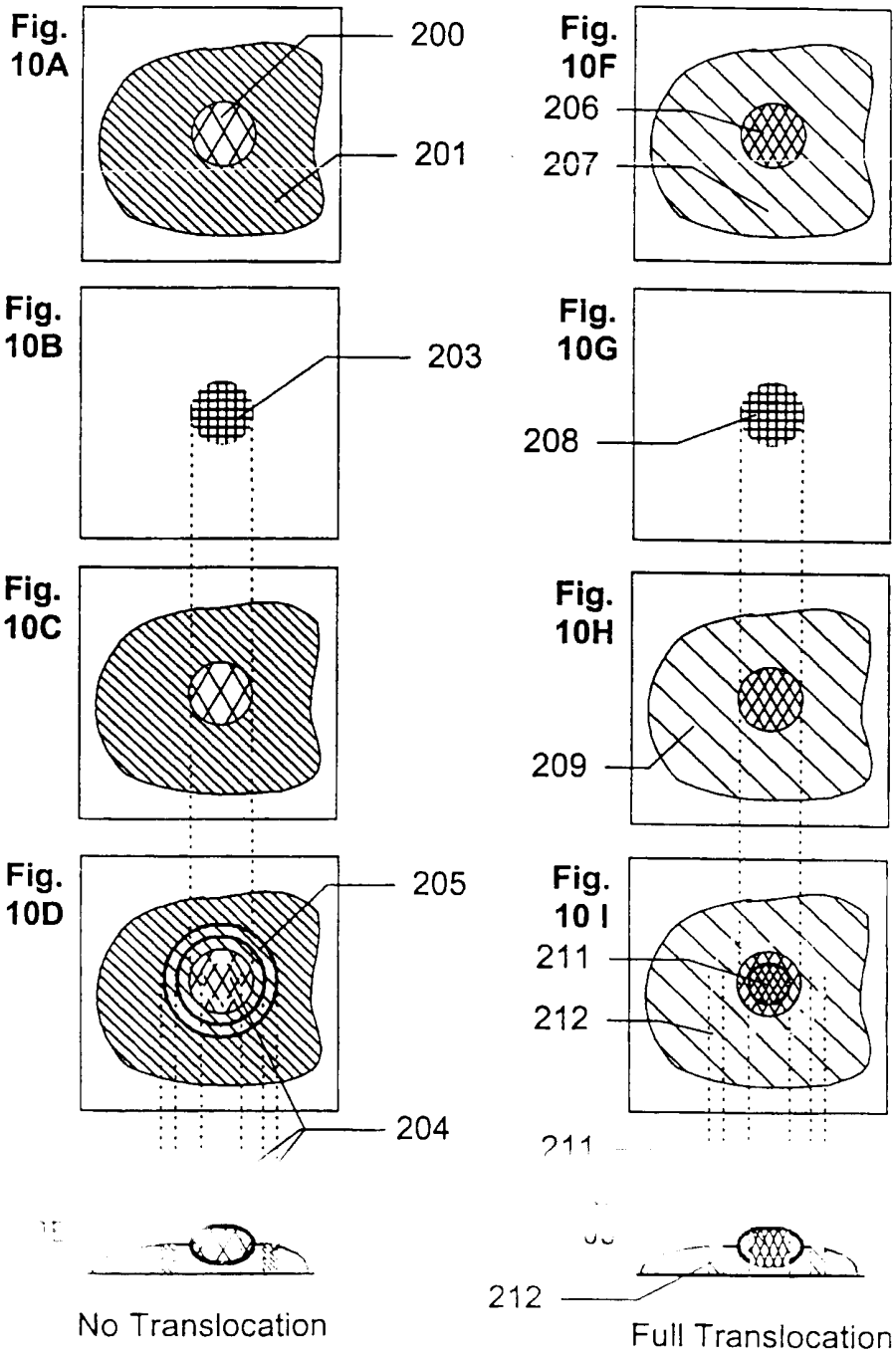


Fig. 10A
200
201
Fig. 10B
203
Fig. 10C
Fig. 10D
205
204
No Translocation
Fig. 10F
206
207
Fig. 10G
208
Fig. 10H
209
Fig. 10 I
211
212
211
212
Full Translocation

Figure 11

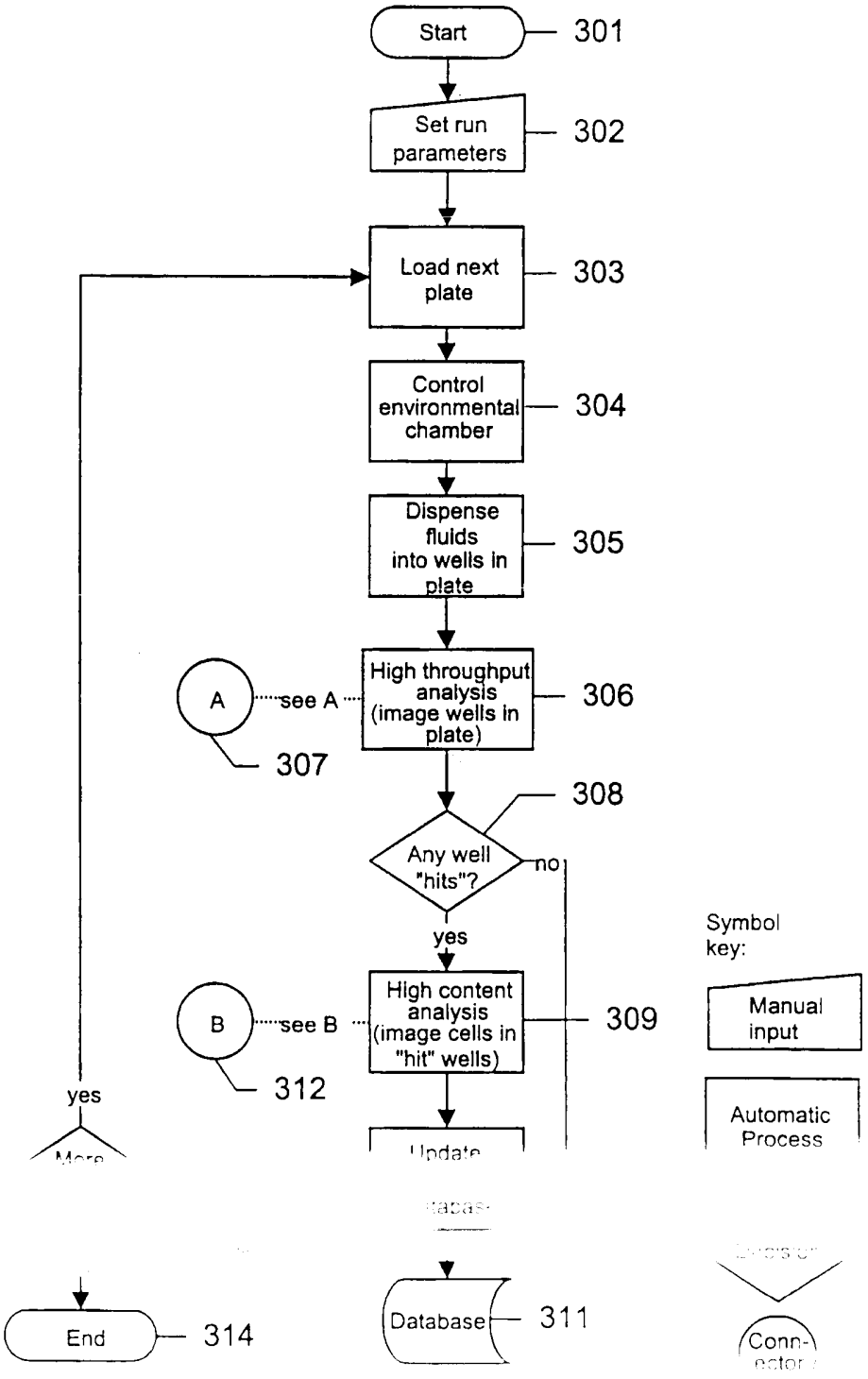

Start
301
Set run parameters
302
Load next plate
303
Control environmental chamber
304
Dispense fluids into wells in plate
305
A
see A
High throughput analysis (image wells in plate)
306
307
308
Any well "hits"?
no
yes
B
see B
High content analysis (image cells in "hit" wells)
309
312
yes
Update
Database
311
End
314
Symbol key:
Manual input
Automatic Process
Conn-ector

Figure 12

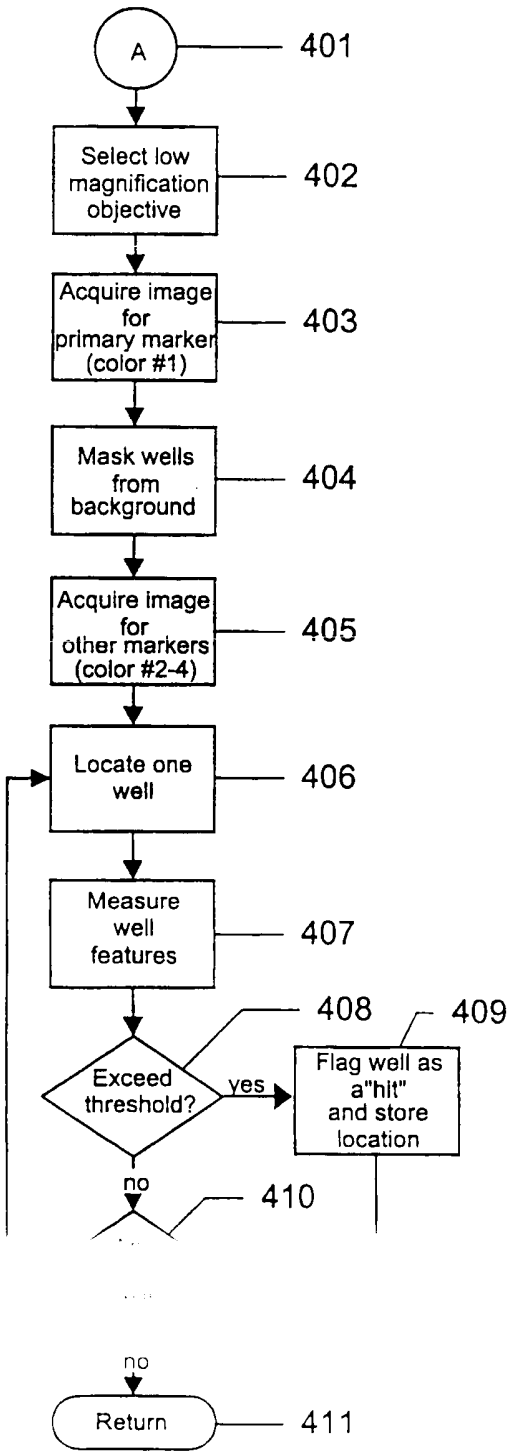

A
401
Select low magnification objective
402
Acquire image for primary marker (color #1)
403
Mask wells from background
404
Acquire image for other markers (color #2-4)
405
Locate one well
406
Measure well features
407
408
409
Exceed threshold?
yes
Flag well as a"hit" and store location
no
410
no
Return
411

Figure 13

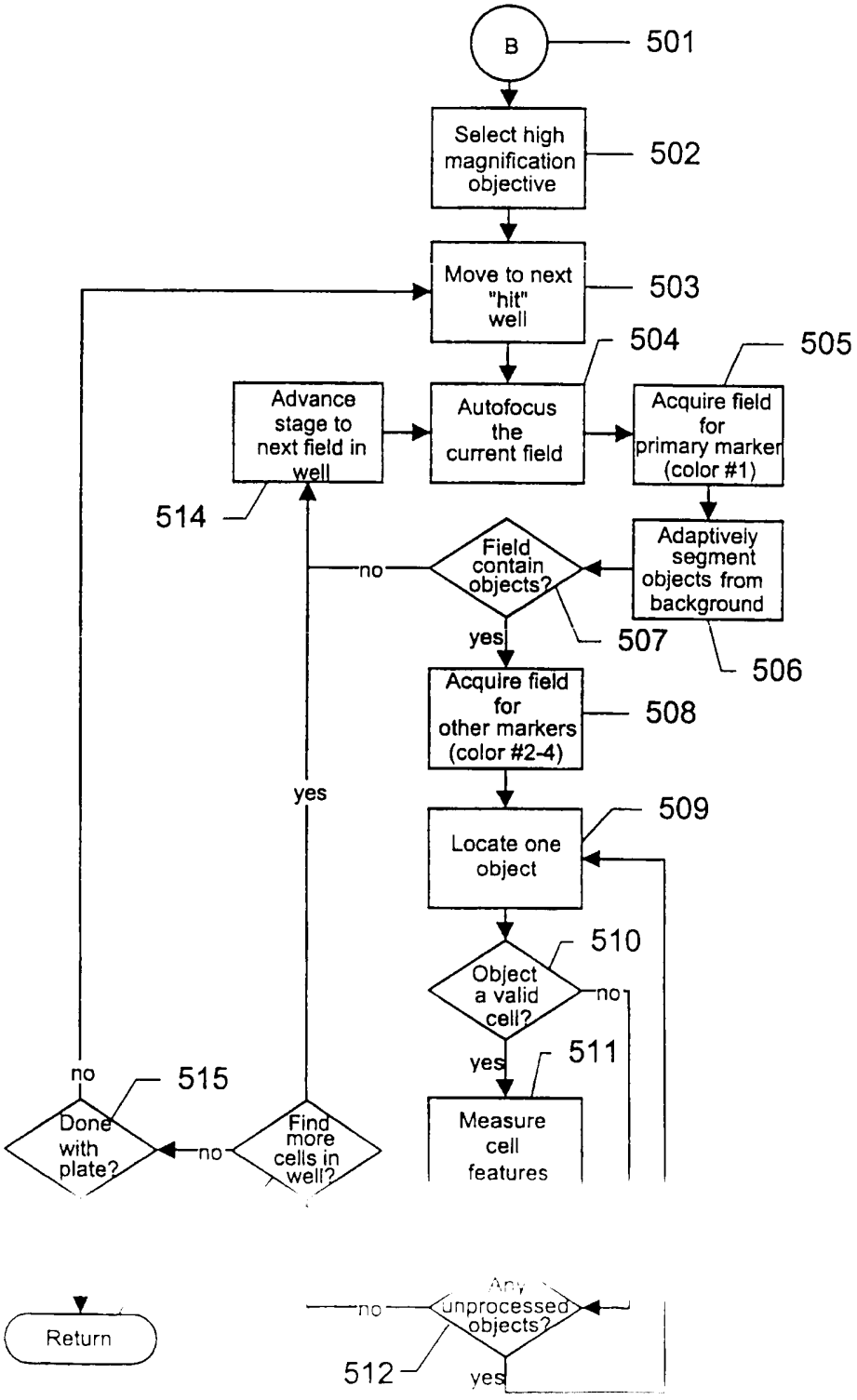

B
501
Select high magnification objective
502
Move to next "hit" well
503
504
505
Advance stage to next field in well
Autofocus the current field
Acquire field for primary marker (color #1)
514
Field contain objects?
no
Adaptively segment objects from background
yes
507
506
Acquire field for other markers (color #2-4)
508
yes
509
Locate one object
510
Object a valid cell?
no
yes
511
no
515
Done with plate?
no
Find more cells in well?
Measure cell features
Return
no
Any unprocessed objects?
512
yes

Figure 14

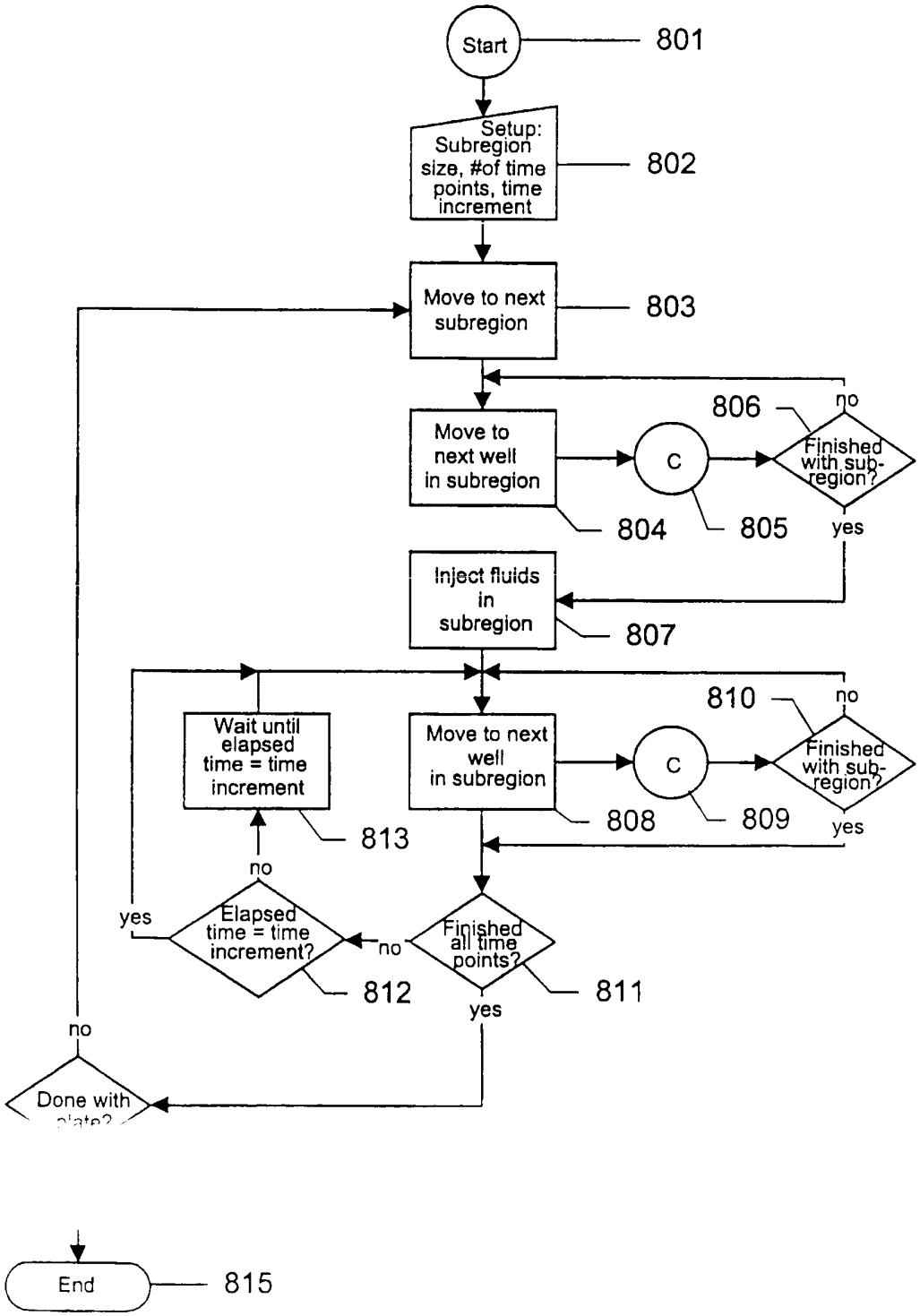

Start
801
Setup: Subregion size, #of time points, time increment
802
Move to next subregion
803
Move to next well in subregion
C
806
no
Finished with sub-region?
yes
804
805
Inject fluids in subregion
807
Wait until elapsed time = time increment
Move to next well in subregion
C
810
no
Finished with sub-region?
yes
808
809
813
no
yes
Elapsed time = time increment?
no
Finished all time points?
812
811
yes
no
Done with
End
815

Figure 15

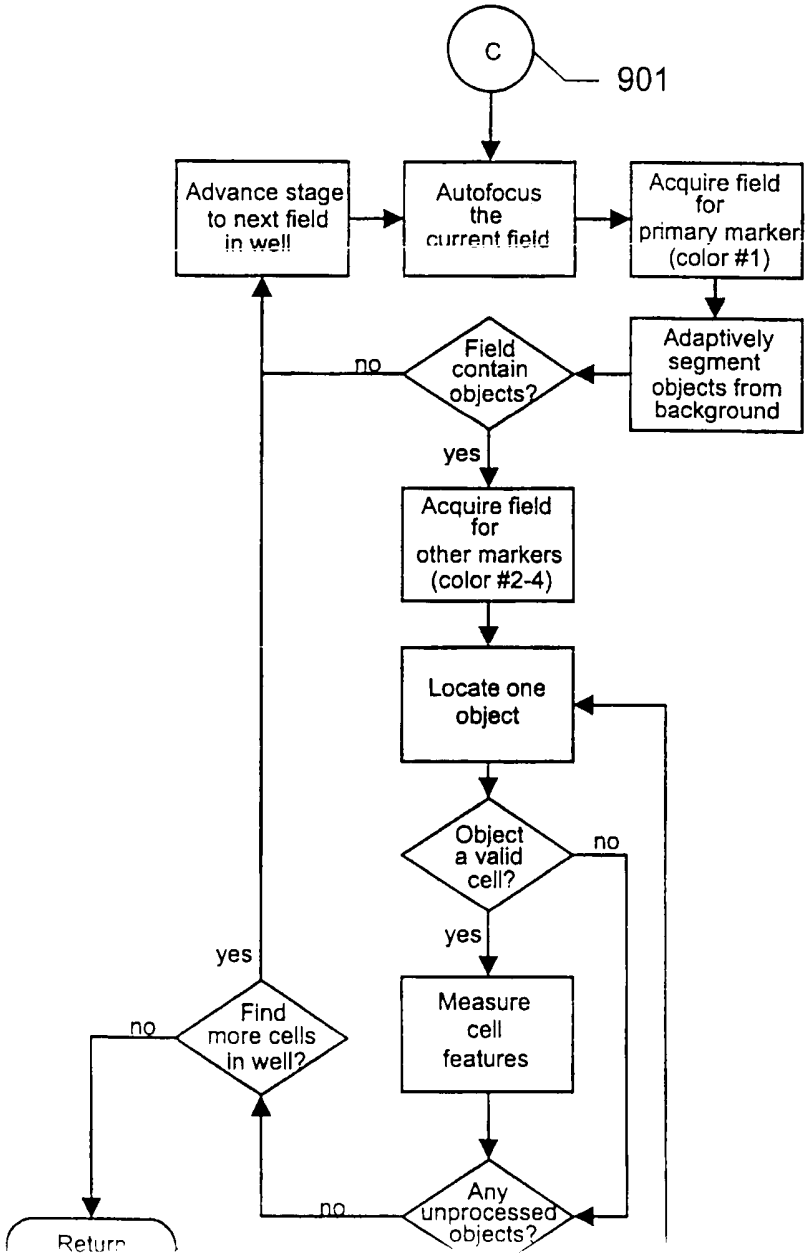

C
901
Advance stage to next field in well
Autofocus the current field
Acquire field for primary marker (color #1)
Adaptively segment objects from background
Field contain objects?
no
yes
Acquire field for other markers (color #2-4)
Locate one object
Object a valid cell?
no
yes
Measure cell features
Find more cells in well?
yes
no
Any unprocessed objects?
no
Return

Figure 16

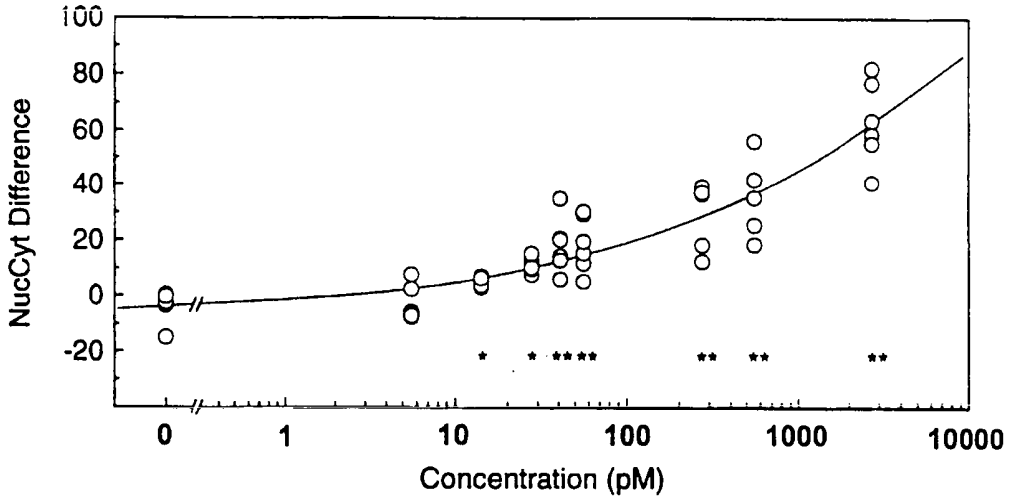

NucCyt Difference
100
80
60
40
20
0
-20
0
1
10
100
1000
10000
Concentration (pM)

Figure 17

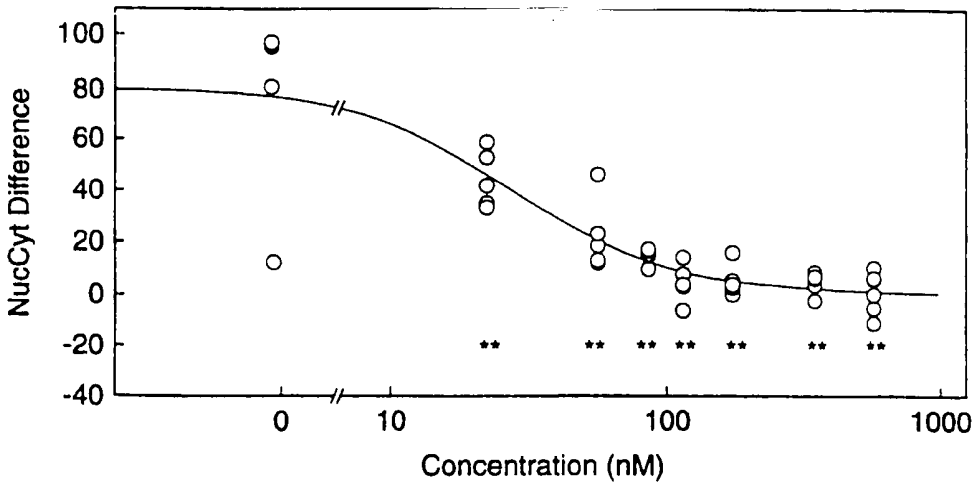

NucCyt Difference
100
80
60
40
20
0
-20
-40
0
10
100
1000
Concentration (nM)
**

Figure 18

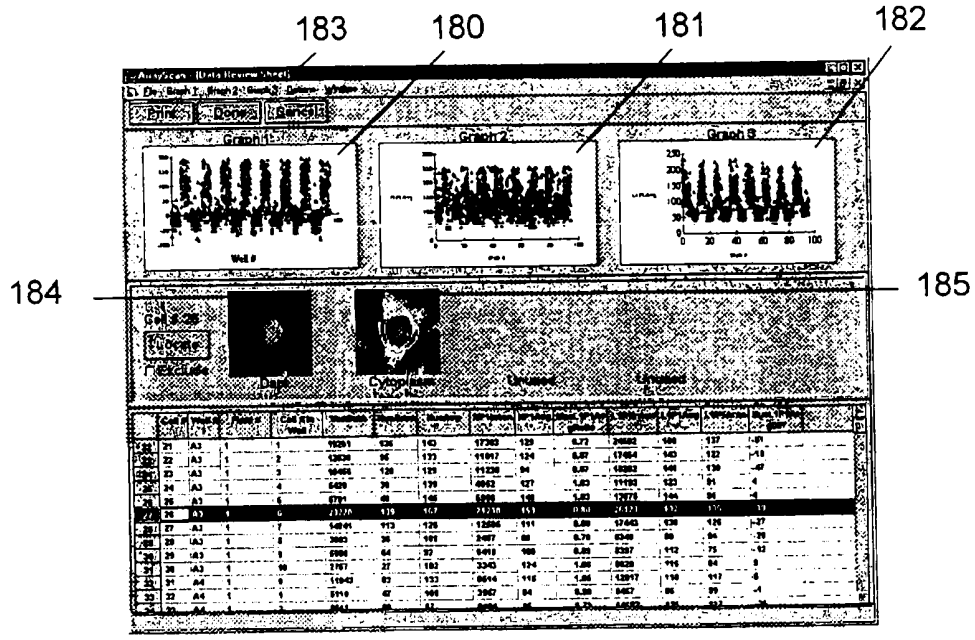


Figure 19

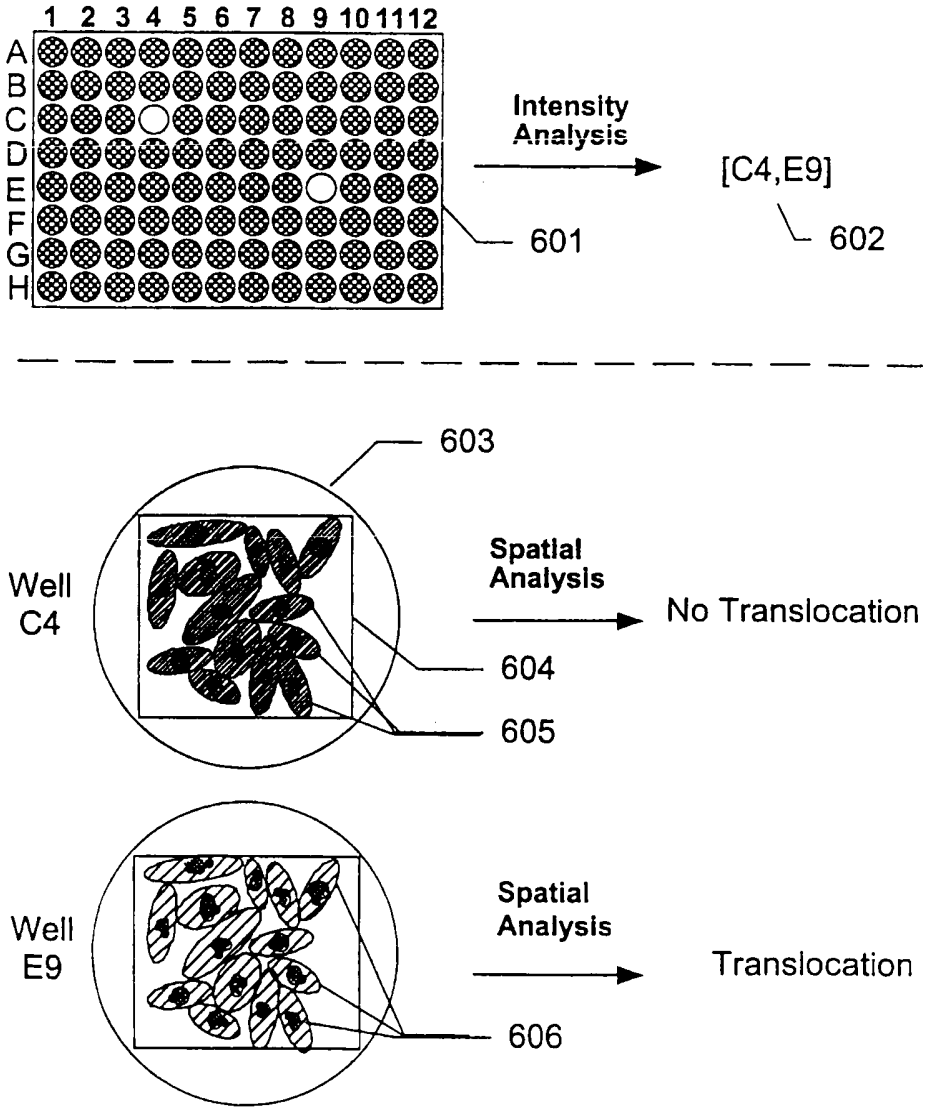

1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
Intensity Analysis
[C4,E9]
601
602
603
Well C4
Spatial Analysis
No Translocation
604
605
Well E9
Spatial Analysis
Translocation
606

Figure 20

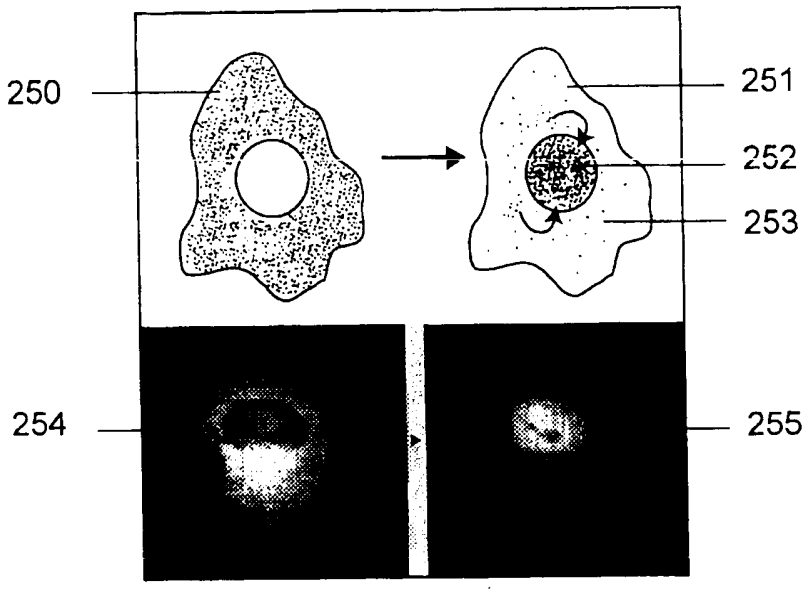


Fig. 20A Fig. 20B

Figure 21

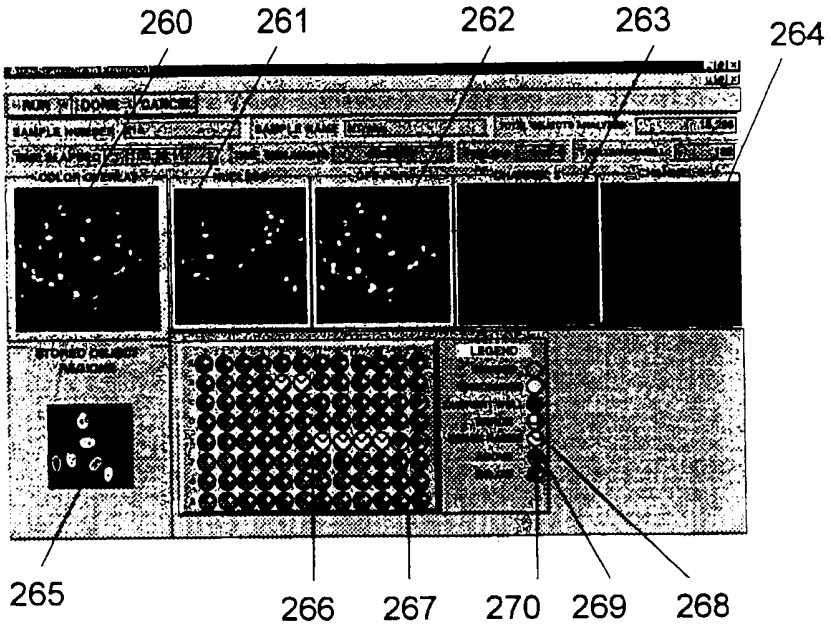


Figure 22

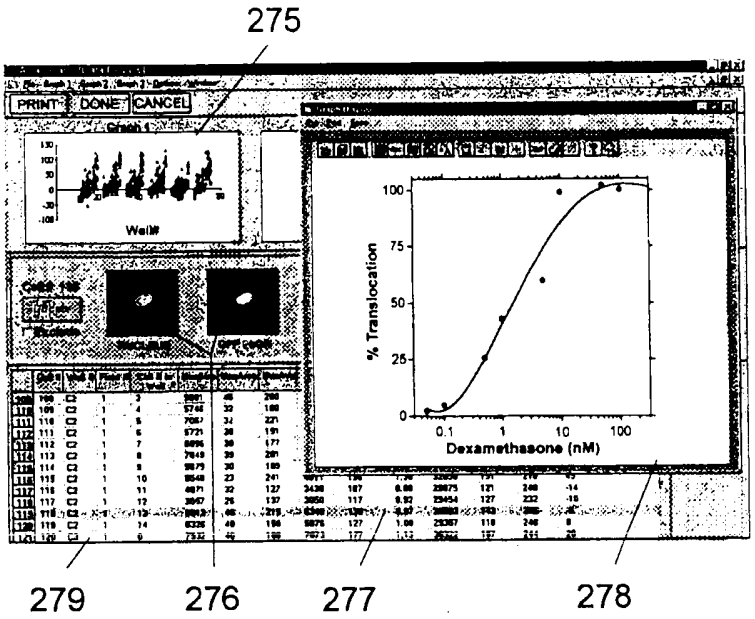

275
PRINT
DONE
CANCEL
Well#
% Translocation
100
75
50
25
0
0.1
1
10
100
Dexamethasone (nM)
279
276
277
278

Figure 23

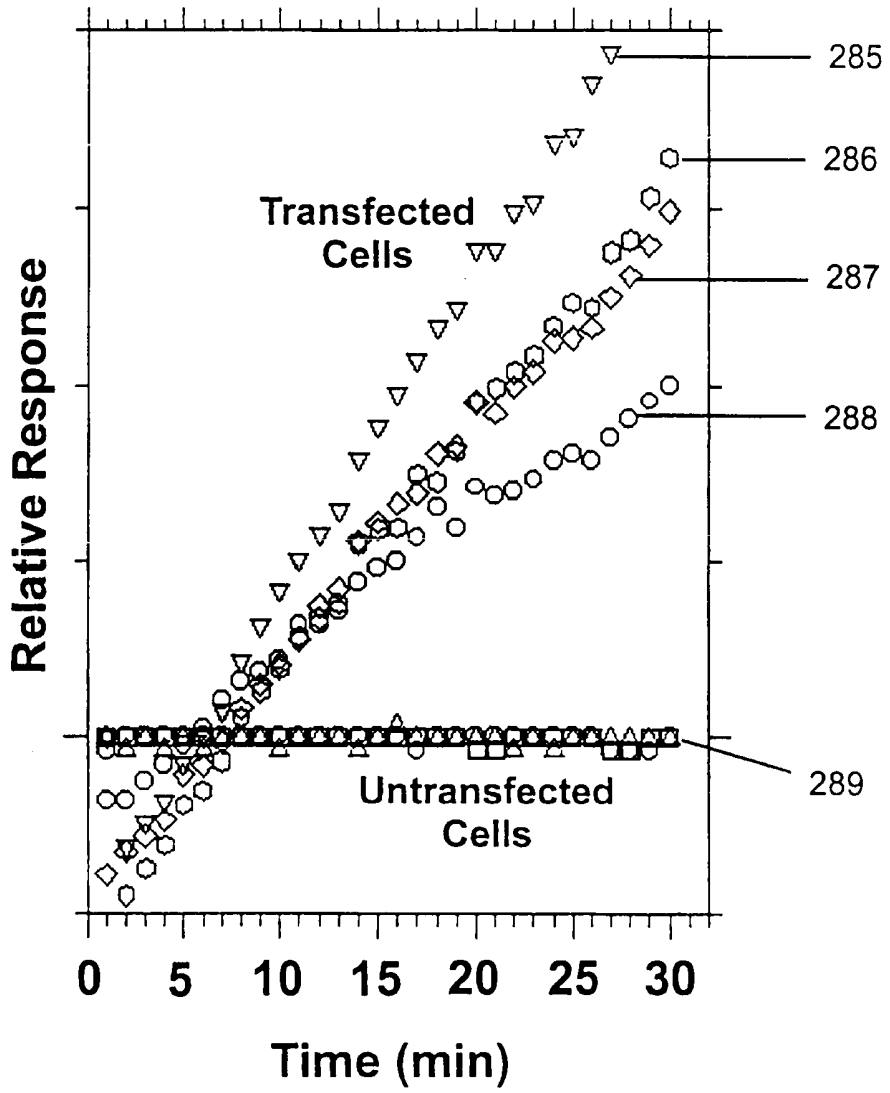

285
286
287
288
289
Transfected Cells
Untransfected Cells
Relative Response
0
5
10
15
20
25
30
Time (min)

Figure 24

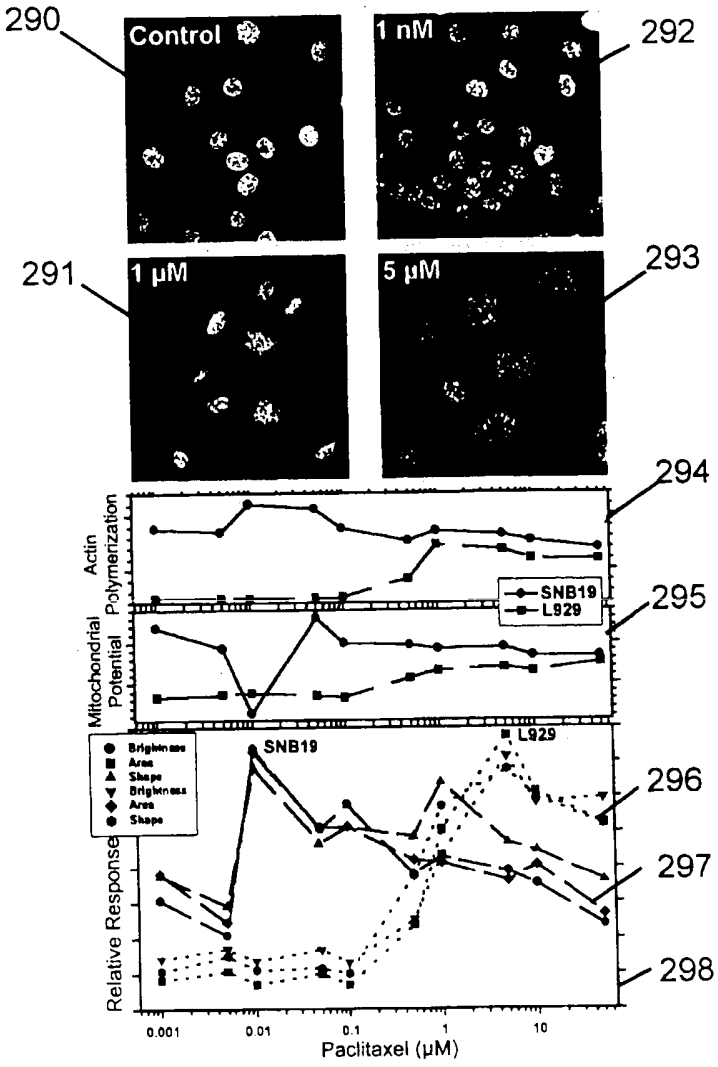

290
Control
1 nM
292
291
1 μM
5 μM
293
294
Actin Polymerization
SNB19
L929
295
Mitochondrial Potential
Brightness
Area
Shape
Brightness
Area
Shape
SNB19
L929
296
297
298
Relative Response
0.001
0.01
0.1
1
10
Paclitaxel (μM)

Figure 25

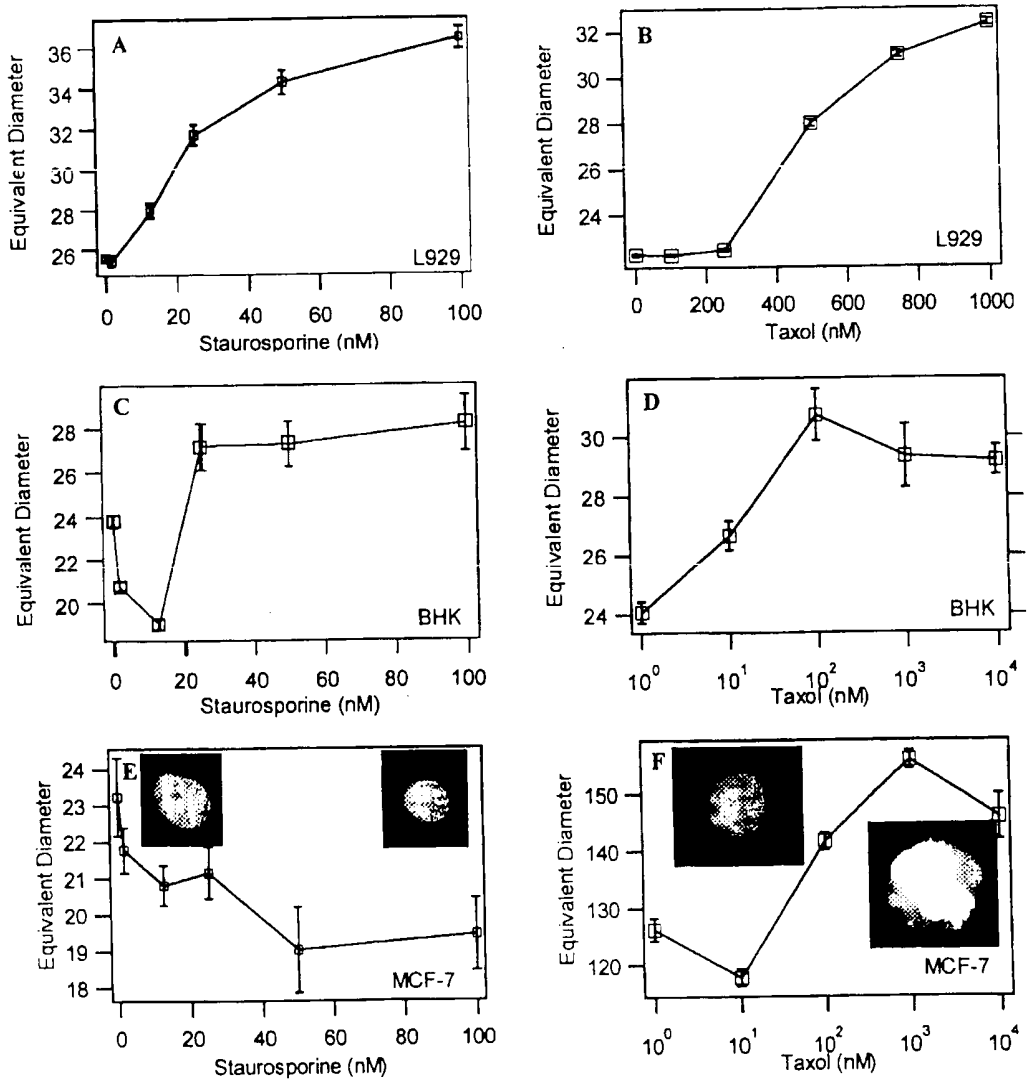

A
Equivalent Diameter
L929
Staurosporine (nM)
B
Equivalent Diameter
L929
Taxol (nM)
C
Equivalent Diameter
BHK
Staurosporine (nM)
D
Equivalent Diameter
BHK
Taxol (nM)
E
Equivalent Diameter
MCF-7
Staurosporine (nM)
F
Equivalent Diameter
MCF-7
Taxol (nM)

Figure 26

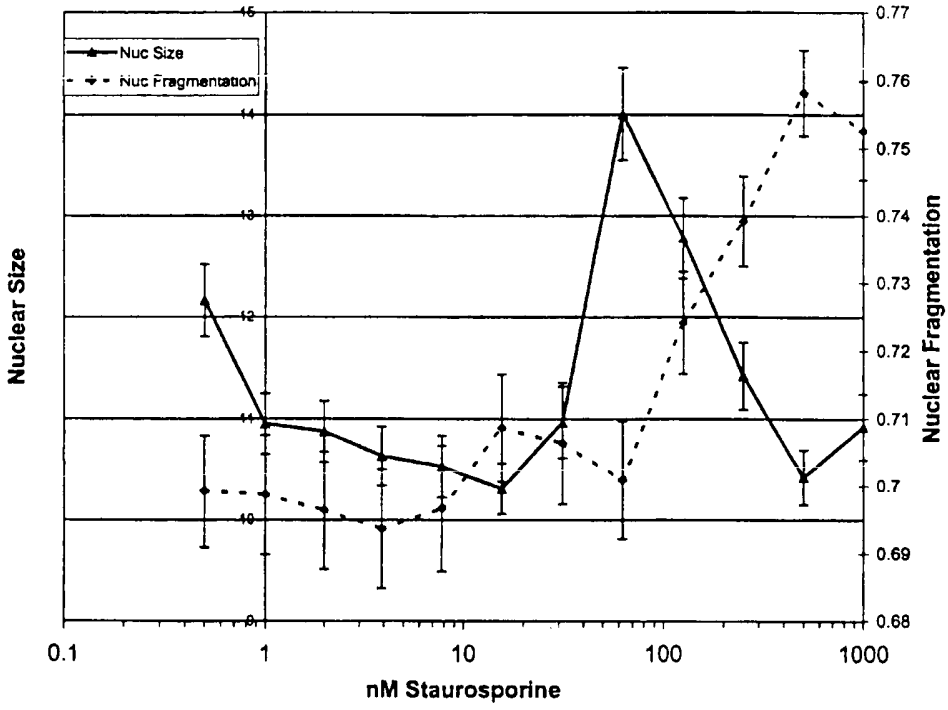

Nuc Size
Nuc Fragmentation
Nuclear Size
Nuclear Fragmentation
nM Staurosporine
0.1
1
10
100
1000
0.77
0.76
0.75
0.74
0.73
0.72
0.71
0.7
0.69
0.68

Figure 27

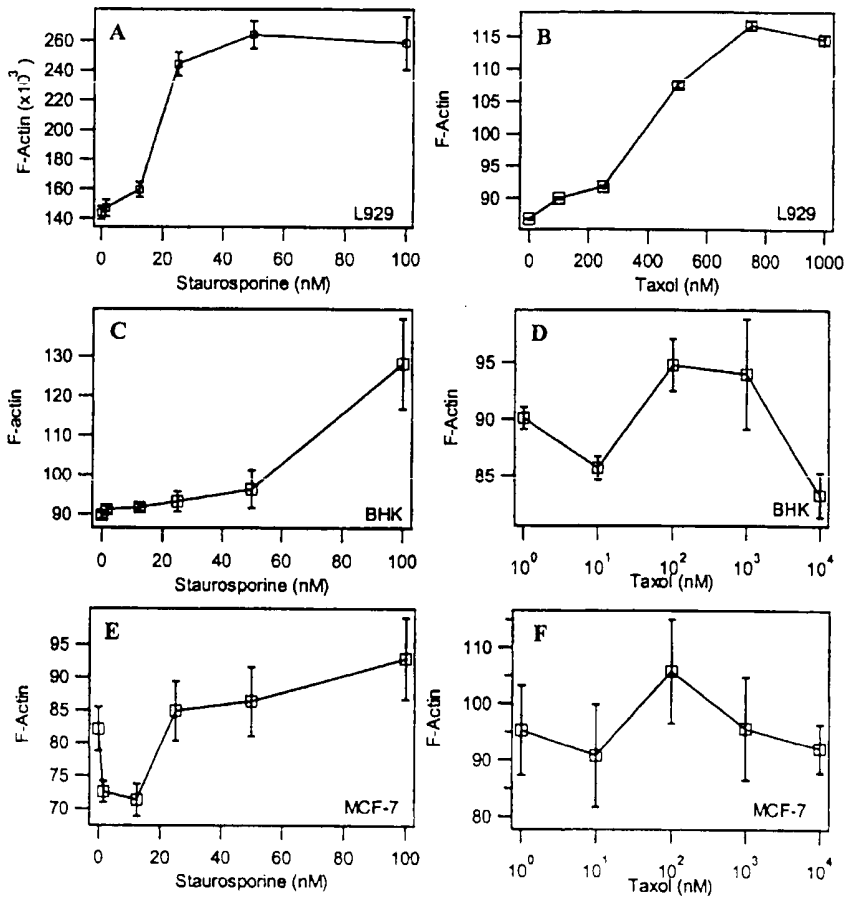

A
F-Actin (x10^3)
L929
Staurosporine (nM)
B
F-Actin
L929
Taxol (nM)
C
F-actin
BHK
Staurosporine (nM)
D
F-Actin
BHK
Taxol (nM)
E
F-Actin
MCF-7
Staurosporine (nM)
F
F-Actin
MCF-7
Taxol (nM)

Figure 28

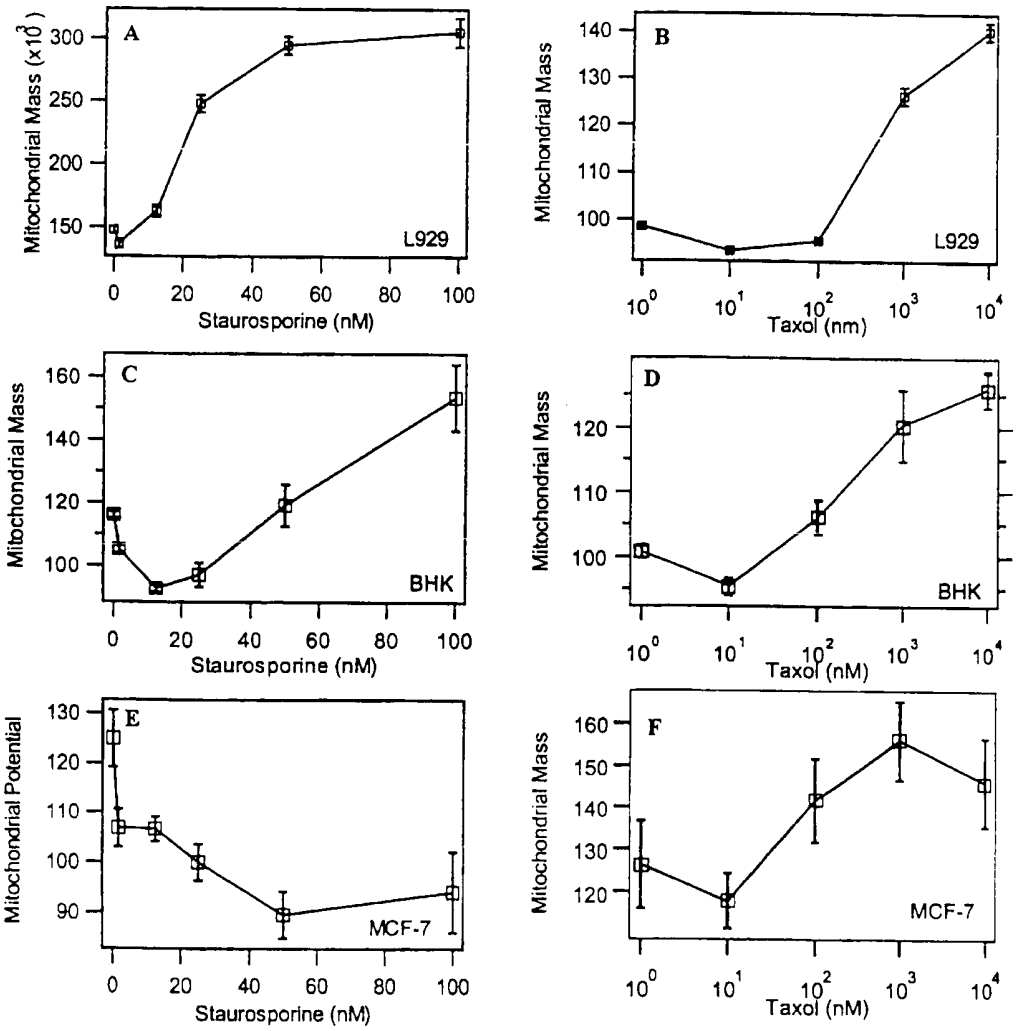

A
Mitochondrial Mass (x10^3)
300
250
200
150
L929
0
20
40
60
80
100
Staurosporine (nM)
B
Mitochondrial Mass
140
130
120
110
100
L929
10^0
10^1
10^2
10^3
10^4
Taxol (nm)
C
Mitochondrial Mass
160
140
120
100
BHK
0
20
40
60
80
100
Staurosporine (nM)
D
Mitochondrial Mass
120
110
100
BHK
10^0
10^1
10^2
10^3
10^4
Taxol (nM)
E
Mitochondrial Potential
130
120
110
100
90
MCF-7
0
20
40
60
80
100
Staurosporine (nM)
F
Mitochondrial Mass
160
150
140
130
120
MCF-7
10^0
10^1
10^2
10^3
10^4
Taxol (nM)

Figure 28G

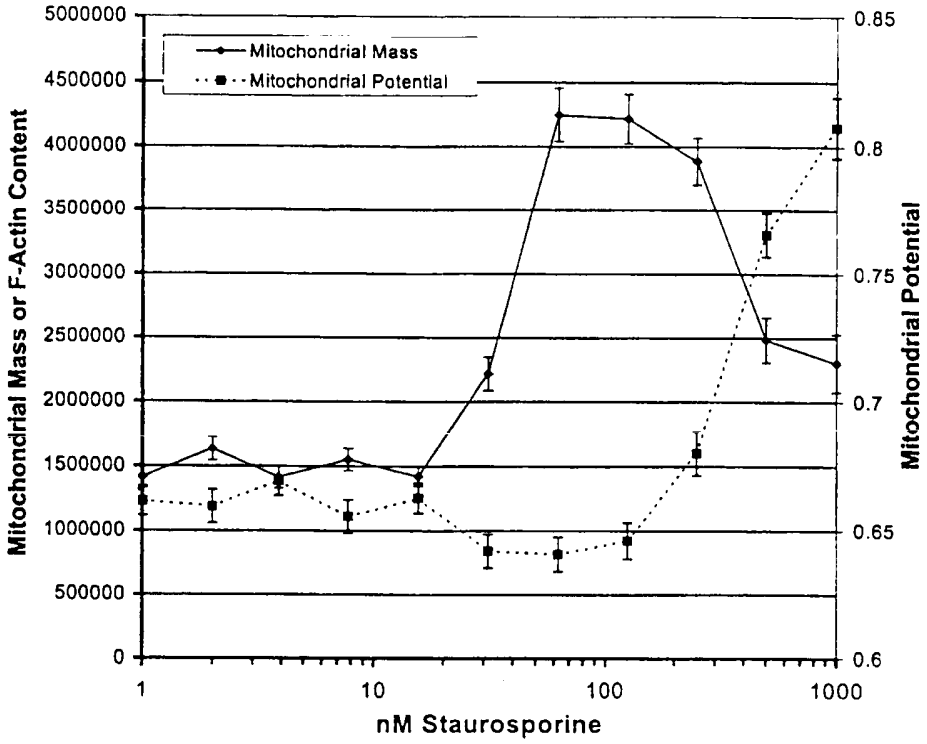

Mitochondrial Mass
Mitochondrial Potential
Mitochondrial Mass or F-Actin Content
Mitochondrial Potential
nM Staurosporine
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
0.85
0.8
0.75
0.7
0.65
0.6
1
10
100
1000

FIGURE 29A

1. SIGNAL SEQUENCES

EPITOPE	SEQUENCE	SEQ ID NO:	REFERENCE
FLAG epitope	5'GACTACAAAGACGACGACGACAAA AA Seq: DYKDDDDK	35 36	Kasir, et al., 1999. J Biol Chem. 274:24873-80.
HA epitope	5'TACCCATACGACGTACCAGACTACGCA AA Seq: YPYDVPDYA	37 38	Smith, et al., 1999. J Biol Chem. 274:19894-900.
KT3 epitope	5'CCACCAGAACCAGAAACA AA seq: PPEPET	39 40	MacArthur and Walter. 1984. J Virol. 52:483-91.
Myc epitope	5'GCAGAAGAACAAAAATTAATAAGCGAAGA AGACTTA AA Seq: AEEQKLISEEDL	41 42	Gosney, et al., 1990. Anticancer Res. 10:623-8.

EYFP: SEQ ID NO: 43 (Nucleic acid); SEQ ID NO:44 (Amino acid)

```
M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC

G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC

G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC

L  V  T  T  F  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K
CTCGTGACCACC TTCGGCTACGGC CTGCAGTGCTTC GCCCGCTACCCC GACCACATGAAG

Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC

F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG

V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC

K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
AAGCTGGAGTAC AACTACAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC

G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
```

```
Y  L  S  Y  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
TACCTGAGCTAC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC

L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAG
```

EGFP: SEQ ID NO:45 (Nucleic acid); SEQ ID NO:46 (Amino acid)

```
M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC

G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC

G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC

L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K
CTCGTGACCACC CTGACCTACGGC GTGCAGTGCTTC AGCCGCTACCCC GACCACATGAAG

Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC

F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG

V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC

K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
AAGCTGGAGTAC AACTACAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC

G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
GGCATCAAGGTG AACTTCAAGATC CGCCACAACATC GAGGACGGCAGC GTGCAGCTCGCC

D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC

Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
TACCTGAGCACC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC

L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAG
```

EBFP: SEQ ID NO:47 (Nucleic acid); SEQ ID NO:48 (Amino acid)

```
G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC

G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC

L  V  T  T  L  T  H  G  V  Q  C  F  S  R  Y  P  D  H  M  K
CTCGTGACCACC CTGACCCACGGC GTGCAGTGCTTC AGCCGCTACCCC GACCACATGAAG

Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC

F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG

V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC

K  L  E  Y  N  F  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
AAGCTGGAGTAC AACTTCAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC

G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
GGCATCAAGGTG AACTTCAAGATC CGCCACAACATC GAGGACGGCAGC GTGCAGCTCGCC

D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC

Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
TACCTGAGCACC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC

L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAG
```

ECFP: SEQ ID NO:49 (Nucleic acid); SEQ ID NO:50 (Amino acid)

```
M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC

G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC

G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC

L  V  T  T  L  T  W  G  V  Q  C  F  S  R  Y  P  D  H  M  K
CTCGTGACCACC CTGACCTGGGGC GTGCAGTGCTTC AGCCGCTACCCC GACCACATGAAG
```

```
F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG

V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC

K  L  E  Y  N  Y  I  S  H  N  V  Y  I  T  A  D  K  Q  K  N
AAGCTGGAGTAC AACTACATCAGC CACAACGTCTAT ATCACCGCCGAC AAGCAGAAGAAC
G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
GGCATCAAGGCC AACTTCAAGATC CGCCACAACATC GAGGACGGCAGC GTGCAGCTCGCC

D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC

Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
TACCTGAGCACC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC

L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAG
```

Fred25: SEQ ID NO:51 (Nucleic acid); SEQ ID NO:52 (Amino acid)

```
M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
ATGGCTAGCAAA GGAGAAGAACTC TTCACTGGAGTT GTCCCAATTCTT GTTGAATTAGAT

G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
GGTGATGTTAAC GGCCACAAGTTC TCTGTCAGTGGA GAGGGTGAAGGT GATGCAACATAC

G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
GGAAAACTTACC CTGAAGTTCATC TGCACTACTGGC AAACTGCCTGTT CCATGGCCAACA

L  V  T  T  L  C  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K
CTAGTCACTACT CTGTGCTATGGT GTTCAATGCTTT TCAAGATACCCG GATCATATGAAA

R  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
CGGCATGACTTT TTCAAGAGTGCC ATGCCCGAAGGT TATGTACAGGAA AGGACCATCTTC

F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
TTCAAAGATGAC GGCAACTACAAG ACACGTGCTGAA GTCAAGTTTGAA GGTGATACCCTT

V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
GTTAATAGAATC GAGTTAAAAGGT ATTGACTTCAAG GAAGATGGCAAC ATTCTGGGACAC

K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
AAATTGGAATAC AACTATAACTCA CACAATGTATAC ATCATGGCAGAC AAACAAAAGAAT

G  I  K  V  N  F  K  T  R  H  N  I  E  D  G  S  V  Q  L  A
```

```
GACCATTATCAA CAAAATACTCCA ATTGGCGATGGC CCTGTCCTTTTA CCAGACAACCAT

Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
TACCTGTCCACA CAATCTGCCCTT TCGAAAGATCCC AACGAAAAGAGA GACCACATGGTC

L  L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  N  *
CTTCTTGAGTTT GTAACAGCTGCT GGGATTACACAT GGCATGGATGAA CTGTACAACTAG
```

FIGURE 29B

2. PROTEASE RECOGNITION SITES

Substrate Recognitions Sequences	Source	Recognition Site	SEQ ID NO	Reference
Caspase-1,4,5	peptide library	5'(TGG,TTA)GAACATGACAA Seq:(W,L)EHD/	53 54	Thornberry et al., 1997, J. Biol. Chem. 272:17907
proCaspase-1	peptide library	5'TGGTTTAAAGAC AA Seq: WFKD/	55 56	Thornberry et al., 1997, J. Biol. Chem. 272:17907
Caspase-2	peptide library	5'GACGAACACGAC AA Seq: DEHD/	57 58	Thornberry et al., 1997, J. Biol. Chem. 272:17907
Caspase 3, 7	PARP	5'GACGAAGTTGAC AA Seq: DEVD/	59 60	Beneke, et al., 1997. Biochem Mol Biol Int. 43:755-61; Thornberry et al., 1997, J. Biol. Chem. 272:17907
ProCaspase 3	Caspase-3	5'ATAGAAACAGAC AA Seq: IETD/	61 62	Tewari, M., et al., 1995. Cell. 81:801-9.
ProCaspase-4,5	peptide library	5'TGGGTAAGAGAC AA Seq: WVRD/	63 64	Thornberry, N.A. et al., 1997, J.Biol. Chem. 272, 17907-17911
Caspase 6	Lamin A, peptide library	5'GTAGAAATAGAC AA Seq: VEID/ 5'GTAGAACACGAC AA Seq: VEHD/	65 66 67 68	Nakajima and Sado. 1993. Biochim Biophys Acta. 1171:311-4; Thornberry et al., 1997, J. Biol. Chem. 272:17907
proCaspase 6	Caspase-6	5'ACAGAAGTAGAC AA Seq: TEVD/	69 70	Fernandes-Alnemri, et al., 1994. J Biol Chem. 269:30761-4.
proCaspase-7	peptide library	5'ATACAAGCAGAC AA Seq: IQAD/	71 72	Thornberry, N.A. et al., 1997, J.Biol. Chem. 272, 17907-17911
Caspase 8	peptide library	5'GTAGAAACAGAC AA Seq: VETD/,	73 74	Muzio, M., et al., 1996. Cell. 85:817-27; Fernandes-Alnemri, et al., 1996. Proc Natl Acad Sci U S A. 93:7464-9;Thornberry et al., 1997, J. Biol. Chem. 272:17907
proCaspase-8	Caspase-8	5'TTAGAAACAGAC AA Seq: LETD/	75 76	Muzio, M., et al., 1996. Cell. 85:817-27; Fernandes-Alnemri, et al., 1996. Proc Natl Acad Sci U S A. 93:7464-9;Thornberry et al., 1997, J. Biol. Chem. 272:17907
Caspase 9	peptide library	5'TTAGAACACGAC AA Seq: LEHD/	77 78	Thornberry, N.A. et al., 1997, J.Biol. Chem. 272, 17907-17911
proCaspase 9	Caspase-9	5'TTAGAACACGAC AA Seq: LEHD/	79 80	Thornberry, N.A. et al., 1997, J.Biol. Chem. 272, 17907-17911
HIV protease		5'AGCCAAAATTAC AA Seq: SQNY/ 5'CCAATAGTACAA AA Seq: PIVQ/	81 82 83 84	Matayoshi, et al., 1990. Science. 247:954-8.
Adenovirus endopeptidase		5'ATGTTTGGAGGA AA Seq: MFGG/ 5'GCAAAAAAAAGA AA Seq: AKKR/	85 86 87 88	Weber and Tihanyi. 1994. Methods Enzymol. 244:595-604.
b-Secretase	Amyloid precursor protein	5'GTAAAAATG AA Seq: VKM/ 5'GACGCAGAATTC DAEF/	89 90 91 92	Hardy et al., 1994, in Amyloid Protein Precursor in Development, Aging, and Alzheimer's Disease, ed. C.L. Masters et al., pp. 190-198.
Cathepsin D		5'AAACCAGCATTATTC AA Seq: KPALF	93 94	Dunn, et al., 1998. Adv Exp Med Biol. 436:133-8.
				[illegible] 1995, Wilhelm et al., 1993, Williams and Auld, 1986, Haugland, R., Handbook of fluorescent probes and research Chemicals 7th ed.

		AA Seq: IEPD/	100	Chem. 272:17907
Anthrax protease	MEK1	5'ATGCCCAAGAAGAAGCCGAC GCCCATCCAGCTGAAC AA Seq: MPKKKPTPIQLN	101 102	Vitale et al., (1998) Biochem Biophys Res Commun 248 (3), 706-711
Anthrax protease	MEK2	5'ATGCTGGCCCGGAGGAAGCCG GTGCTGCCGGCGCTCACCATCA AC AA Seq: MLARRKPVLPALTIN	103 104	Vitale et al., (1998) Biochem Biophys Res Commun 248 (3), 706-711
tetanus/botulinum	cellubrevin	5'GCCTCGCAGTTTGAAACA AA Seq: ASQFET	105 106	McMahon et al., Nature 364:346-349; Martin et al., J. Cell Biol. In press
tetanus/botulinum	synaptobrevin/ VAMP3	5'GCTTCTCAATTTGAAACG AA Seq: ASQFET	107 108	Schiavo et al., (1992) Nature 359, 832-5
Botulinum neurotoxin A	SNAP-25	5'GCCAACCAACGTGCAACA AA Seq: ANQ/RAT	109 110	Zhao, et al. Gene 145 (2), 313-314 (1994)
Botulinum neurotoxin B	VAMP	5'GCTTCTCAATTTGAAACG AA Seq: ASQ/FET	111 112	
Botulinum neurotoxin C	Syntaxin	5'ACGAAAAAAGCTGTGAAA AA Seq: TKK/AVK	113 114	Martin et al., J. Leukoc. Biol. 65 (3), 397-406 (1999)
Botulinum neurotoxin D	VAMP	5'GACCAGAAGCTCTCTGAG AA Seq: DQK/LSE	115 116	
Botulinum neurotoxin E	SNAP-25	5'ATCGACAGGATCATGGAG AA Seq: IDR/IME	117 118	
Botulinum neurotoxin F	VAMP	5'AGAGACCAGAAGCTCTCT AA Seq: RDQ/KLS	119 120	
Botulinum neurotoxin G	VAMP	5'ACGAGCGCAGCCAAGTTG AA Seq: TSA/AKL	121 122	

FIGURE 29C

3. PRODUCT/REACTANT TARGET SEQUENCES

Target	Target Source	Target domain (Product or Reactant)	SEQ ID NO	Reference
Cytoplasm/cytoskeleton	Annexin II	5'ATGTCTACTGTCCACGAAATCCTGTGCAAG CTCAGCTTGGAGGGTGTTCATTCTACACCCCC AAGTGCC 3'	123	Eberhard, et al., 1997, Mol. Biol. Cell 8:293a.
		(Amino acid seq: M S T V H E I L C K L S L E G V H S T P P S A)	124	
Inner surface of plasma membrane	farnesylation	5'ATGGGATCTACATTAAGCGCAGAAGACAAA GCAGCAGTAGAAAGAAGCAAAATGATAGAC AGAAACTTATTAAGAGAAGACGGAGAAAAA GCTGCTAGA3'	125	Ferruccio G, et al., J. Biol. Chem. 274, 5843-5850, 1999
		(AA seq: M G C T L S A E D K A A V E R S K M I D R N L R E D G E K A A R	126	
Nucleus	NFkB p50	5'AGAAGGAAACGACAAAAG	127	Henkel, T et al., Cell 68, 1121-1133, 1992
		(AA seq: R R K R Q K)	128	
Nucleolus	NOLP	5'AGAAAACGTATACGTACTTACCTCAAGTCC TGCAGGCGGATGAAAAGAAGTGGTTTTGAGA TGTCTCGACCTATTCCTTCCCACCTTACT	129	Ueki, et al., 1998. Biochem Biophys Res Commun. 252:97-102.
		(AA seq: R K R I R T Y L K S C R R M K R S G F E M S R P I P S H L T)	130	
Mitochondria	cytochrome c oxidase	5'ATGTCCGTCCTGACGCCGCTGCTGCTGCGG GGCTTGACAGGCTCGGCCCGGCGGCTCCCAG TGCCGCGCGCCAAGATCCATTCGTTG	131	Rizzuto, et al., 1989. J Biol Chem. 264:10595-600.
		(AA Seq: M S V L T P L L L R G L T G S A R R L P V P R A L I H S L)	132	
Nuclear Envelope	ODV-E66 & ODV-E25	5'ATGAGCATTGTTTTAATAATTGTTATTTGGA TTTTTTTAATATGTTTTTTATATTTAAGCAACA GCAAAGATCCCAGAGTACCAGTTGAATTAAT G	133	Hong, T, et al. PNAS, 94, 4050-4055, 1997
		(AA Seq: M S I V L I I V I V V I F L I C F L Y L S N S K D P R V P V E L M)	134	
Golgi	Calreticulin	5'ATGAGGCTTCGGGAGCCGCTCCTGAGCGGC AGCGCCGCGATGCCAGGCGCGTCCCTACAGC GGGCCTGCCGCCTGCTCGTGGCCGTCTGCGCT CTGCACCTTGGCGTCACCCTCGTTTACTACCT GGCTGGCCGCGACCTGAGCCGCCTGCCCCAA CTGGTCGGAGTCTCCACACCGCTGCAGGGCG GCTCGAACAGTGCCGCCGCCATCGGGCAGTC CTCCGGGGAGCTCCGGACCGGAGGGGCC	135	Fliegel, L., et al., J. Biol. Chem. 264, 21522-21528, 1989.
		(AA Seq: M R L R E P L L S G S A A M P G A S L Q R A C R L L V A V C A L H L G V T L V Y Y L A G R D L S R L P Q L V G V S T P L Q G G S N S A A A I G Q S S G E L R T G G A)	136	
Endoplasmic reticulum	D-AKAP1	5'GAAACAATAAGACCTATAAGAAGATGTAGT ACATTTACATCTACAGACAGCAAAAUGGCAA TTCAATTAAGATCTCCCTTTCCATTAGCATTA CCAGGAATGTTAGCTTTATTAGGATGGTGGT GGTTTTTCAGTAGAAAAAAA	137	Huang, LJ. Et al., J. Cell. Biol. 145, 951-959, 1999

		(AA SEQ:A L Q K K L E E L E L D E	140	272, 51, 32642-32648
Size exclusion	PROJ domain of MAP4	5'GCCGACCTCAGTCTTGTGGATGCGTTGACA GAACCACCTCCAGAAATTGAGGGAGAAATAA AGCGAGACTTCATGGCTGCGCTGGAGGCAGA GCCCTATGATGACATCGTGGGAGAAACTGTG GAGAAAACTGAGTTTATTCCTCTCCTGGATGG TGATGAGAAAACCGGGAACTCAGAGTCCAAA AAGAAACCCTGCTTAGACACTAGCCAGGTTG AAGGTATCCCATCTTCTAAACCAACACTCCTA GCCAATGGTGATCATGGAATGGAGGGGAATA ACACTGCAGGGTCTCCAACTGACTTCCTTGAA GAGAGAGTGGACTATCCGGATTATCAGAGCA GCCAGAACTGGCCAGAAGATGCAAGCTTTTG TTTCCAGCCTCAGCAAGTGTTAGATACTGACC AGGCTGAGCCCTTTAACGAGCACCGTGATGA TGGTTTGGCAGATCTGCTCTTTGTCTCCAGTG GACCCACGAACGCTTCTGCATTTACAGAGCG AGACAATCCTTCAGAAGACAGTTACGGTATG CTTCCCTGTGACTCATTTGCTTCCACGGCTGT TGTATCTCAGGAGTGGTCTGTGGGAGCCCCA AACTCTCCATGTTCAGAGTCCTGTGTCTCCCC AGAGGTTACTATAGAAACCCTACAGCCAGCA ACAGAGCTCTCCAAGGCAGCAGAAGTGGAAT CAGTGAAAGAGCAGCTGCCAGCTAAAGCATT GGAAACGATGGCAGAGCAGACCACTGATGTG GTGCACTCTCCATCCACAGACACAACACCAG GCCCAGACACAGAGGCAGCACTGGCTAAAGA CATAGAAGAGATCACCAAGCCAGATGTGATA TTGGCAAATGTCACGCAGCCATCTACTGAAT CGGATATGTTCCTGGCCCAGGACATGGAACT ACTCACAGGAACAGAGGCAGCCCACGCTAAC AATATCATATTGCCTACAGAACCAGACGAAT CTTCAACCAAGGATGTAGCACCACCTATGGA AGAAGAAATTGTCCCAGGCAATGATA	141	West, (1991). J Biol Chem 266(32): 21886-96; Olson, K. R. (1995). J Cell Biol 130(3): 639-50.
		(AA SEQ: A D L S L V D A L T E P P P E I E G E I K R D F M A A L E A E P Y D D I V G E T V E K T E F I P L L D G D E K T G N S E S K K K P C L D T S Q V E G I P S S K P T L L A N G D H G M E G N N T A G S P T D F L E E R V D Y P D Y Q S S Q N W P E D A S F C F Q P Q Q V L D T D Q A E P F N E H R D D G L A D L L F V S S G P T N A S A F T E R D N P S E D S Y G M L P C D S F A S T A V V S Q E W S V G A P N S P C S E S C V S P E V T I E T L Q P A T E L S K A A E V E S V K E Q L P A K A L E T M A E Q T T D V V H S P S T D T T P G P D T E A A L A K D I E E I T K P D V I L A N V T Q P S T E S D M F L A Q D M E L L T G T E A A H A N N I I L P T E P D E S S T K D V A P P M E E E I V P G N D T T S P K E T E T T L P I K M D L A P P E D V L L T K E T E L A P A K G M V S L S E I E E A L A K N D V R S A E I P V A Q E T V V S E T E V V L A T E V V L P S D P I T T L T K D V T L P L E A E R P L V T D M T P S L E T E M T L G K E T A P P T E T N L G M A K D M S P L P E S E V T L G K D V V I L P E T K V A E F N N V T P L S E E E V T S V K D M S P S A E T E A P L A K N A D L H S G T E L I V D N S M A P A S D L A L P L E T K V A T V P I K D K G	142	
Vesicle membrane	Synaptobrevin	5' ATGTGGGCAATCGGGATTACTGTTCT GGTTATCTTCATCATCATCATCATCGTG TGGGTTGTC	143	Schiavo et al., (1992) Nature 359,832-5

Vesicle membrane	Cellubrevin	5'ATGTGGGCGATAGGGATCAGTGTCCT GGTGATCATTGTCATCATCATCATCGTG TGGTGTG (AA SEQ: M W A I G I S V L V I I V I I I I V W C)	145 146	McMahon et al., Nature 364:346-349; Martin et al., J. Cell Biol. In press
Nuclear Export	MEK2	5'GACCTGCAGAAGAAGCTGGAGGAGCT GGAACTTGACGAG AA SEQ: D L Q K K L E E L E L D E	147 148	Zheng and Guan, J. Biol. Chem. 268:11435-11439. 1993
Peroxisome	PX	5'TCTAAACTG AA SEQ: S K L	149 150	Amery et al., Biochem. J. 336:367-371 (1998)

Microtubules (MAP4) SEQ ID NO:151 (Nucleic acid); SEQ ID NO:152 (amino acid)

MAP4 :

```
M  A  D  L  S  L  V  D  A  L  T  E  P  P  P  E  I  E  G  E
ATGGCCGACCTC AGTCTTGTGGAT GCGTTGACAGAA CCACCTCCAGAA ATTGAGGGAGAA
TACCGGCTGGAG TCAGAACACCTA CGCAACTGTCTT GGTGGAGGTCTT TAACTCCCTCTT

I  K  R  D  F  M  A  A  L  E  A  E  P  Y  D  D  I  V  G  E
ATAAAGCGAGAC TTCATGGCTGCG CTGGAGGCAGAG CCCTATGATGAC ATCGTGGGAGAA
TATTTCGCTCTG AAGTACCGACGC GACCTCCGTCTC GGGATACTACTG TAGCACCCTCTT

T  V  E  K  T  E  F  I  P  L  L  D  G  D  E  K  T  G  N  S
ACTGTGGAGAAA ACTGAGTTTATT CCTCTCCTGGAT GGTGATGAGAAA ACCGGGAACTCA
TGACACCTCTTT TGACTCAAATAA GGAGAGGACCTA CCACTACTCTTT TGGCCCTTGAGT

E  S  K  K  K  P  C  L  D  T  S  Q  V  E  G  I  P  S  S  K
GAGTCCAAAAAG AAACCCTGCTTA GACACTAGCCAG GTTGAAGGTATC CCATCTTCTAAA
CTCAGGTTTTTC TTTGGGACGAAT CTGTGATCGGTC CAACTTCCATAG GGTAGAAGATTT

P  T  L  L  A  N  G  D  H  G  M  E  G  N  N  T  A  G  S  P
CCAACACTCCTA GCCAATGGTGAT CATGGAATGGAG GGGAATAACACT GCAGGGTCTCCA
GGTTGTGAGGAT CGGTTACCACTA GTACCTTACCTC CCCTTATTGTGA CGTCCCAGAGGT

T  D  F  L  E  E  R  V  D  Y  P  D  Y  Q  S  S  Q  N  W  P
ACTGACTTCCTT GAAGAGAGAGTG GACTATCCGGAT TATCAGAGCAGC CAGAACTGGCCA
TGACTGAAGGAA CTTCTCTCTCAC CTGATAGGCCTA ATAGTCTCGTCG GTCTTGACCGGT

E  D  A  S  F  C  F  Q  P  Q  Q  V  L  D  T  D  Q  A  E  P
GAAGATGCAAGC TTTTGTTTCCAG CCTCAGCAAGTG TTAGATACTGAC CAGGCTGAGCCC
CTTCTACGTTCG AAAACAAAGGTC GGAGTCGTTCAC AATCTATGACTG GTCCGACTCGGG

F  N  E  H  R  D  D  G  L  A  D  L  L  F  V  S  S  G  P  T
TTTAACGAGCAC CGTGATGATGGT TTGGCAGATCTG CTCTTTGTCTCC AGTGGACCCACG
AAATTGCTCGTG GCACTACTACCA AACCGTCTAGAC GAGAAACAGAGG TCACCTGGGTGC
```

```
C  D  S  F  A  S  T  A  V  V  S  Q  E  W  S  V  G  A  P  N
TGTGACTCATTT GCTTCCACGGCT GTTGTATCTCAG GAGTGGTCTGTG GGAGCCCCAAAC
ACACTGAGTAAA CGAAGGTGCCGA CAACATAGAGTC CTCACCAGACAC CCTCGGGGTTTG

S  P  C  S  E  S  C  V  S  P  E  V  T  I  E  T  L  Q  P  A
TCTCCATGTTCA GAGTCCTGTGTC TCCCCAGAGGTT ACTATAGAAACC CTACAGCCAGCA
AGAGGTACAAGT CTCAGGACACAG AGGGGTCTCCAA TGATATCTTTGG GATGTCGGTCGT

T  E  L  S  K  A  A  E  V  E  S  V  K  E  Q  L  P  A  K  A
ACAGAGCTCTCC AAGGCAGCAGAA GTGGAATCAGTG AAAGAGCAGCTG CCAGCTAAAGCA
TGTCTCGAGAGG TTCCGTCGTCTT CACCTTAGTCAC TTTCTCGTCGAC GGTCGATTTCGT

L  E  T  M  A  E  Q  T  T  D  V  V  H  S  P  S  T  D  T  T
TTGGAAACGATG GCAGAGCAGACC ACTGATGTGGTG CACTCTCCATCC ACAGACACAACA
AACCTTTGCTAC CGTCTCGTCTGG TGACTACACCAC GTGAGAGGTAGG TGTCTGTGTTGT

P  G  P  D  T  E  A  A  L  A  K  D  I  E  E  I  T  K  P  D
CCAGGCCCAGAC ACAGAGGCAGCA CTGGCTAAAGAC ATAGAAGAGATC ACCAAGCCAGAT
GGTCCGGGTCTG TGTCTCCGTCGT GACCGATTTCTG TATCTTCTCTAG TGGTTCGGTCTA

V  I  L  A  N  V  T  Q  P  S  T  E  S  D  M  F  L  A  Q  D
GTGATATTGGCA AATGTCACGCAG CCATCTACTGAA TCGGATATGTTC CTGGCCCAGGAC
CACTATAACCGT TTACAGTGCGTC GGTAGATGACTT AGCCTATACAAG GACCGGGTCCTG

M  E  L  L  T  G  T  E  A  A  H  A  N  N  I  I  L  P  T  E
ATGGAACTACTC ACAGGAACAGAG GCAGCCCACGCT AACAATATCATA TTGCCTACAGAA
TACCTTGATGAG TGTCCTTGTCTC CGTCGGGTGCGA TTGTTATAGTAT AACGGATGTCTT

P  D  E  S  S  T  K  D  V  A  P  P  M  E  E  E  I  V  P  G
CCAGACGAATCT TCAACCAAGGAT GTAGCACCACCT ATGGAAGAAGAA ATTGTCCCAGGC
GGTCTGCTTAGA AGTTGGTTCCTA CATCGTGGTGGA TACCTTCTTCTT TAACAGGGTCCG

N  D  T  T  S  P  K  E  T  E  T  T  L  P  I  K  M  D  L  A
AATGATACGACA TCCCCCAAAGAA ACAGAGACAACA CTTCCAATAAAA ATGGACTTGGCA
TTACTATGCTGT AGGGGGTTTCTT TGTCTCTGTTGT GAAGGTTATTTT TACCTGAACCGT

P  P  E  D  V  L  L  T  K  E  T  E  L  A  P  A  K  G  M  V
CCACCTGAGGAT GTGTTACTTACC AAAGAAACAGAA CTAGCCCCAGCC AAGGGCATGGTT
GGTGGACTCCTA CACAATGAATGG TTTCTTTGTCTT GATCGGGGTCGG TTCCCGTACCAA

S  L  S  E  I  E  E  A  L  A  K  N  D  V  R  S  A  E  I  P
TCACTCTCAGAA ATAGAAGAGGCT CTGGCAAAGAAT GATGTTCGCTCT GCAGAAATACCT
AGTGAGAGTCTT TATCTTCTCCGA GACCGTTTCTTA CTACAAGCGAGA CGTCTTTATGGA

V  A  Q  E  T  V  V  S  E  T  E  V  V  L  A  T  E  V  V  L
GTGGCTCAGGAG ACAGTGGTCTCA GAAACAGAGGTG GTCCTGGCAACA GAAGTGGTACTG
CACCGAGTCCTC TGTCACCAGAGT CTTTGTCTCCAC CAGGACCGTTGT CTTCACCATGAC

P  S  D  P  I  T  T  L  T  K  D  V  T  L  P  L  E  A  E  R
CCCTCAGATCCC ATAACAACATTG ACAAAGGATGTG ACACTCCCTTA GAAGCAGAGCG
```

```
P  L  V  T  D  M  T  P  S  L  E  T  E  M  T  L  G  K  E  T
CCGTTGGTGACG GACATGACTCCA TCTCTGGAAACA GAAATGACCCTA GGCAAAGAGACA
GGCAACCACTGC CTGTACTGAGGT AGAGACCTTTGT CTTTACTGGGAT CCGTTTCTCTGT

A  P  P  T  E  T  N  L  G  M  A  K  D  M  S  P  L  P  E  S
GCTCCACCCACA GAAACAAATTTG GGCATGGCCAAA GACATGTCTCCA CTCCCAGAATCA
CGAGGTGGGTGT CTTTGTTTAAAC CCGTACCGGTTT CTGTACAGAGGT GAGGGTCTTAGT

E  V  T  L  G  K  D  V  V  I  L  P  E  T  K  V  A  E  F  N
GAAGTGACTCTG GGCAAGGACGTG GTTATACTTCCA GAAACAAAGGTG GCTGAGTTTAAC
CTTCACTGAGAC CCGTTCCTGCAC CAATATGAAGGT CTTTGTTTCCAC CGACTCAAATTG

N  V  T  P  L  S  E  E  E  V  T  S  V  K  D  M  S  P  S  A
AATGTGACTCCA CTTTCAGAAGAA GAGGTAACCTCA GTCAAGGACATG TCTCCGTCTGCA
TTACACTGAGGT GAAAGTCTTCTT CTCCATTGGAGT CAGTTCCTGTAC AGAGGCAGACGT

E  T  E  A  P  L  A  K  N  A  D  L  H  S  G  T  E  L  I  V
GAAACAGAGGCT CCCCTGGCTAAG AATGCTGATCTG CACTCAGGAACA GAGCTGATTGTG
CTTTGTCTCCGA GGGGACCGATTC TTACGACTAGAC GTGAGTCCTTGT CTCGACTAACAC

D  N  S  M  A  P  A  S  D  L  A  L  P  L  E  T  K  V  A  T
GACAACAGCATG GCTCCAGCCTCC GATCTTGCACTG CCCTTGGAAACA AAAGTAGCAACA
CTGTTGTCGTAC CGAGGTCGGAGG CTAGAACGTGAC GGGAACCTTTGT TTTCATCGTTGT

V  P  I  K  D  K  G  T  V  Q  T  E  E  K  P  R  E  D  S  Q
GTTCCAATTAAA GACAAAGGAACT GTACAGACTGAA GAAAAACCACGT GAAGACTCCCAG
CAAGGTTAATTT CTGTTTCCTTGA CATGTCTGACTT CTTTTTGGTGCA CTTCTGAGGGTC

L  A  S  M  Q  H  K  G  Q  S  T  V  P  P  C  T  A  S  P  E
TTAGCATCTATG CAGCACAAGGGA CAGTCAACAGTA CCTCCTTGCACG GCTTCACCAGAA
AATCGTAGATAC GTCGTGTTCCCT GTCAGTTGTCAT GGAGGAACGTGC CGAAGTGGTCTT

P  V  K  A  A  E  Q  M  S  T  L  P  I  D  A  P  S  P  L  E
CCAGTCAAAGCT GCAGAACAAATG TCTACCTTACCA ATAGATGCACCT TCTCCATTAGAG
GGTCAGTTTCGA CGTCTTGTTTAC AGATGGAATGGT TATCTACGTGGA AGAGGTAATCTC

N  L  E  Q  K  E  T  P  G  S  Q  P  S  E  P  C  S  G  V  S
AACTTAGAGCAG AAGGAAACGCCT GGCAGCCAGCCT TCTGAGCCTTGC TCAGGAGTATCC
TTGAATCTCGTC TTCCTTTGCGGA CCGTCGGTCGGA AGACTCGGAACG AGTCCTCATAGG

R  Q  E  E  A  K  A  A  V  G  V  T  G  N  D  I  T  T  P  P
CGGCAAGAAGAA GCAAAGGCTGCT GTAGGTGTGACT GGAAATGACATC ACTACCCCGCCA
GCCGTTCTTCTT CGTTTCCGACGA CATCCACACTGA CCTTTACTGTAG TGATGGGGCGGT

N  K  E  P  P  P  S  P  E  K  K  A  K  P  L  A  T  T  Q  P
AACAAGGAGCCA CCACCAAGCCCA GAAAAGAAAGCA AAGCCTTTGGCC ACCACTCAACCT
TTGTTCCTCGGT GGTGGTTCGGGT CTTTTCTTTCGT TTCGGAAACCGG TGGTGAGTTGGA

A  K  T  S  T  S  K  A  K  T  Q  P  T  S  L  P  K  Q  P  A
```

```
P  T  T  S  G  G  L  N  K  K  P  M  S  L  A  S  G  S  V  P
CCCACCACCTCT GGTGGGTTGAAT AAAAAACCCATG AGCCTCGCCTCA GGCTCAGTGCCA
GGGTGGTGGAGA CCACCCAACTTA TTTTTTGGGTAC TCGGAGCGGAGT CCGAGTCACGGT

A  A  P  H  K  R  P  A  A  A  T  A  T  A  R  P  S  T  L  P
GCTGCCCCACAC AAACGCCCTGCT GCTGCCACTGCT ACTGCCAGGCCT TCCACCCTACCT
CGACGGGGTGTG TTTGCGGGACGA CGACGGTGACGA TGACGGTCCGGA AGGTGGGATGGA

A  R  D  V  K  P  K  P  I  T  E  A  K  V  A  E  K  R  T  S
GCCAGAGACGTG AAGCCAAAGCCA ATTACAGAAGCT AAGGTTGCCGAA AAGCGGACCTCT
CGGTCTCTGCAC TTCGGTTTCGGT TAATGTCTTCGA TTCCAACGGCTT TTCGCCTGGAGA

P  S  K  P  S  S  A  P  A  L  K  P  G  P  K  T  T  P  T  V
CCATCCAAGCCT TCATCTGCCCCA GCCCTCAAACCT GGACCTAAAACC ACCCCAACCGTT
GGTAGGTTCGGA AGTAGACGGGGT CGGGAGTTTGGA CCTGGATTTTGG TGGGGTTGGCAA

S  K  A  T  S  P  S  T  L  V  S  T  G  P  S  S  R  S  P  A
TCAAAAGCCACA TCTCCCTCAACT CTTGTTTCCACT GGACCAAGTAGT AGAAGTCCAGCT
AGTTTTCGGTGT AGAGGGAGTTGA GAACAAAGGTGA CCTGGTTCATCA TCTTCAGGTCGA

T  T  L  P  K  R  P  T  S  I  K  T  E  G  K  P  A  D  V  K
ACAACTCTGCCT AAGAGGCCAACC AGCATCAAGACT GAGGGGAAACCT GCTGATGTCAAA
TGTTGAGACGGA TTCTCCGGTTGG TCGTAGTTCTGA CTCCCCTTTGGA CGACTACAGTTT

R  M  T  A  K  S  A  S  A  D  L  S  R  S  K  T  T  S  A  S
AGGATGACTGCT AAGTCTGCCTCA GCTGACTTGAGT CGCTCAAAGACC ACCTCTGCCAGT
TCCTACTGACGA TTCAGACGGAGT CGACTGAACTCA GCGAGTTTCTGG TGGAGACGGTCA

S  V  K  R  N  T  T  P  T  G  A  A  P  P  A  G  M  T  S  T
TCTGTGAAGAGA AACACCACTCCC ACTGGGGCAGCA CCCCCAGCAGGG ATGACTTCCACT
AGACACTTCTCT TTGTGGTGAGGG TGACCCCGTCGT GGGGGTCGTCCC TACTGAAGGTGA

R  V  K  P  M  S  A  P  S  R  S  S  G  A  L  S  V  D  K  K
CGAGTCAAGCCC ATGTCTGCACCT AGCCGCTCTTCT GGGGCTCTTTCT GTGGACAAGAAG
GCTCAGTTCGGG TACAGACGTGGA TCGGCGAGAAGA CCCCGAGAAAGA CACCTGTTCTTC

P  T  S  T  K  P  S  S  S  A  P  R  V  S  R  L  A  T  T  V
CCCACTTCCACT AAGCCTAGCTCC TCTGCTCCCAGG GTGAGCCGCCTG GCCACAACTGTT
GGGTGAAGGTGA TTCGGATCGAGG AGACGAGGGTCC CACTCGGCGGAC CGGTGTTGACAA

S  A  P  D  L  K  S  V  R  S  K  V  G  S  T  E  N  I  K  H
TCTGCCCCTGAC CTGAAGAGTGTT CGCTCCAAGGTC GGCTCTACAGAA AACATCAAACAC
AGACGGGGACTG GACTTCTCACAA GCGAGGTTCCAG CCGAGATGTCTT TTGTAGTTTGTG

Q  P  G  G  G  R  A  K  V  E  K  K  T  E  A  A  T  T  A  G
CAGCCTGGAGGA GGCCGGGCCAAA GTAGAGAAAAAA ACAGAGGCAGCT ACCACAGCTGGG
GTCGGACCTCCT CCGGCCCGGTTT CATCTCTTTTTT TGTCTCCGTCGA TGGTGTCGACCC

K  P  E  P  N  A  V  T  K  A  A  G  S  I  A  S  A  Q  K  P
```

```
GGACGACCCTTT CAGGTCTATCAT AGGTTTTTTCAC TCGATGTCAGTA TAAGTTAGGTTC

C  V  S  K  D  N  I  K  H  V  P  G  C  G  N  V  Q  I  Q  N
TGTGTTTCCAAG GACAATATTAAG CATGTCCCTGGA TGTGGCAATGTT CAGATTCAGAAC
ACACAAAGGTTC CTGTTATAATTC GTACAGGGACCT ACACCGTTACAA GTCTAAGTCTTG

K  K  V  D  I  S  K  V  S  S  K  C  G  S  K  A  N  I  K  H
AAGAAAGTGGAC ATATCCAAGGTC TCCTCCAAGTGT GGGTCCAAAGCT AATATCAAGCAC
TTCTTTCACCTG TATAGGTTCCAG AGGAGGTTCACA CCCAGGTTTCGA TTATAGTTCGTG

K  P  G  G  G  D  V  K  I  E  S  Q  K  L  N  F  K  E  K  A
AAGCCTGGTGGA GGAGATGTCAAG ATTGAAAGTCAG AAGTTGAACTTC AAGGAGAAGGCC
TTCGGACCACCT CCTCTACAGTTC TAACTTTCAGTC TTCAACTTGAAG TTCCTCTTCCGG

Q  A  K  V  G  S  L  D  N  V  G  H  F  P  A  G  G  A  V  K
CAAGCCAAAGTG GGATCCCTTGAT AACGTTGGCCAC TTTCCTGCAGGA GGTGCCGTGAAG
GTTCGGTTTCAC CCTAGGGAACTA TTGCAACCGGTG AAAGGACGTCCT CCACGGCACTTC

T  E  G  G  G  S  E  A  L  P  C  P  G  P  P  A  G  E  E  P
ACTGAGGGCGGT GGCAGTGAGGCC CTTCCGTGTCCA GGCCCCCCCGCT GGGGAGGAGCCA
TGACTCCCGCCA CCGTCACTCCGG GAAGGCACAGGT CCGGGGGGGCGA CCCCTCCTCGGT

V  I  P  E  A  A  P  D  R  G  A  P  T  S  A  S  G  L  S  G
GTCATCCCTGAG GCTGCGCCTGAC CGTGGCGCCCCT ACTTCAGCCAGT GGCCTCAGTGGC
CAGTAGGGACTC CGACGCGGACTG GCACCGCGGGGA TGAAGTCGGTCA CCGGAGTCACCG

H  T  T  L  S  G  G  G  D  Q  R  E  P  Q  T  L  D  S  Q  I
CACACCACCCTG TCAGGGGGTGGT GACCAAAGGGAG CCCCAGACCTTG GACAGCCAGATC
GTGTGGTGGGAC AGTCCCCCACCA CTGGTTTCCCTC GGGGTCTGGAAC CTGTCGGTCTAG

Q  E  T  S  I  *
CAGGAGACAAGC ATCTAA
GTCCTCTGTTCG TAGATT
```

Figure 30

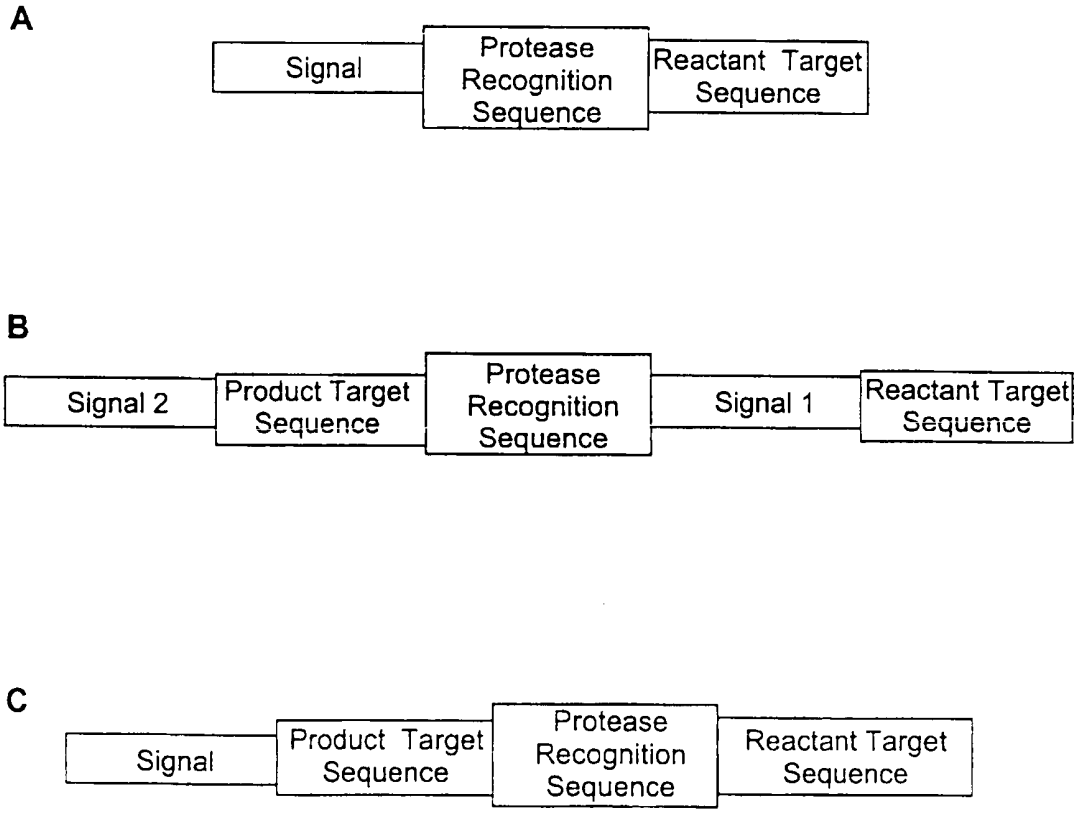
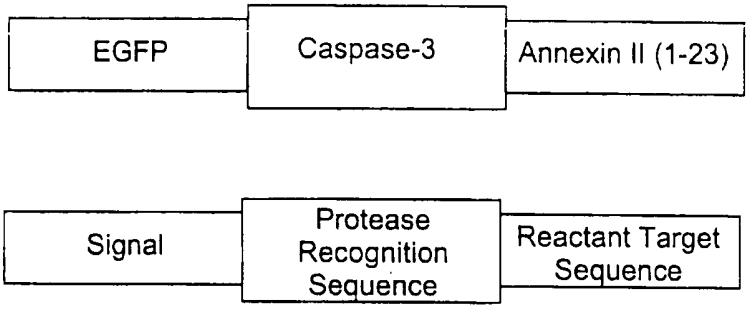

A
Signal
Protease Recognition Sequence
Reactant Target Sequence
B
Signal 2
Product Target Sequence
Protease Recognition Sequence
Signal 1
Reactant Target Sequence
C
Signal
Product Target Sequence
Protease Recognition Sequence
Reactant Target Sequence

Figure 31

A


EGFP
Caspase-3
Annexin II (1-23)
Signal
Protease Recognition Sequence
Reactant Target Sequence

B

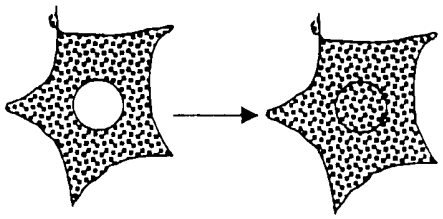


FIGURE 32

SEQ ID 1-2

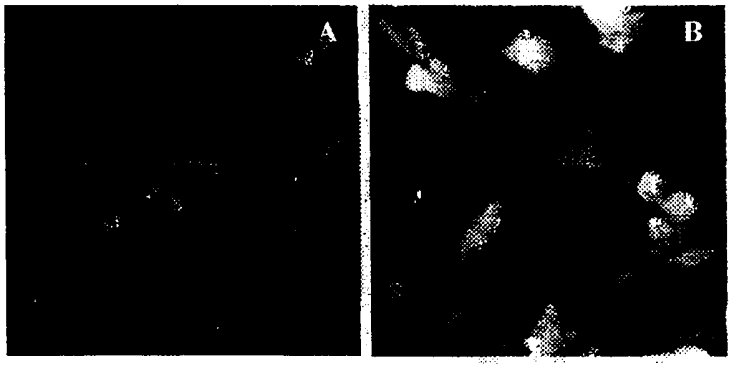
Original caspase biosensor: GFP italic, DEVD bold and Annexin II underlined

```
      +1  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
       1  ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC
          TACCACTCGTTC CCGCTCCTCGAC AAGTGGCCCCAC CACGGGTAGGAC CAGCTCGACCTG

      +1  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
      61  GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC
          CCGCTGCATTTG CCGGTGTTCAAG TCGCACAGGCCG CTCCCGCTCCCG CTACGGTGGATG

      +1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
     121  GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC
          CCGTTCGACTGG GACTTCAAGTAG ACGTGGTGGCCG TTCGACGGGCAC GGGACCGGGTGG

      +1  L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K
     181  CTCGTGACCACC CTGACCTACGGC GTGCAGTGCTTC AGCCGCTACCCC GACCACATGAAG
          GAGCACTGGTGG GACTGGATGCCG CACGTCACGAAG TCGGCGATGGGG CTGGTGTACTTC

      +1  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     241  CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC
          GTCGTGCTGAAG AAGTTCAGGCGG TACGGGCTTCCG ATGCAGGTCCTC GCGTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
     301  TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG
          AAGTTCCTGCTG CCGTTGATGTTC TGGGCGCGGCTC CACTTCAAGCTC CCGCTGTGGGAC

      +1  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
     361  GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC
          CACTTGGCGTAG CTCGACTTCCCG TAGCTGAAGTTC CTCCTGCCGTTG TAGGACCCCGTG

      +1  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
     421  AAGCTGGAGTAC AACTACAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC
          TTCGACCTCATG TTGATGTTGTCG GTGTTGCAGATA TAGTACCGGCTG TTCGTCTTCTTG

      +1  G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
                              BstYI
                             -------
     481  GGCATCAAGGTG AACTTCAAGATC CGCCACAACATC GAGGACGGCAGC GTGCAGCTCGCC
          CCGTAGTTCCAC TTGAAGTTCTAG GCGGTGTTGTAG CTCCTGCCGTCG CACGTCGAGCGG

      +1  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
     541  GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC
          CTGGTGATGGTC GTCTTGTGGGGG TAGCCGCTGCCG GGGCACGACGAC GGGCTGTTGGTG

      +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
                                                                AvaII
                                                               -----
     601  TACCTGAGCACC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC
          ATGGACTCGTGG GTCAGGCGGGAC TCGTTTCTGGGG TTGCTCTTCGCG CTAGTGTACCAG

     661  CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAGTCC
          GACGACCTCAAG CACTGGCGGCGG CCCTAGTGAGAG CCGTACCTGCTC GACATGTTCAGG
```

```
  +1  G  L  R  S  G  A  G  A  G  A  G  A  G  A  G  A  D  E  V  D
             BglII
            ------
             BstYI
            ------
721  GGACTCAGATCT GGCGCCGGCGCT GGAGCCGGAGCT GGCGCCGGAGCC GACGAGGTGGAC
     CCTGAGTCTAGA CCGCGGCCGCGA CCTCGGCCTCGA CCGCGGCCTCGG CTGCTCCACCTG

  +1  G  A  G  A  D  E  V  D  G  A  M  S  T  V  H  E  I  L  C  K
                                        AccI
                                       -------
781  GGCGCCGGCGCC GATGAAGTAGAT GGCGCCATGTCT ACTGTCCACGAA ATCCTGTGCAAG
     CCGCGGCCGCGG CTACTTCATCTA CCGCGGTACAGA TGACAGGTGCTT TAGGACACGTTC

  +1  L  S  L  E  G  D  H  S  T  P  P  S  A  Y  *
841  CTCAGCTTGGAG GGTGATCATTCT ACACCCCCAAGT GCCTATTGA
     GAGTCGAACCTC CCACTAGTAAGA TGTGGGGGTTCA CGGATAACT
```

```
atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggc
cacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctac
cccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttc
ttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag
ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaac
gtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggc
agcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccac
tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtg
accgccgccgggatcactctcggcatggacgagctgtacaagtccggactcagatctggcgccggcgctggagcc
ggagctggcgccggagccgacgaggtggacggcgccggcgccgatgaagtagatggcgccatgtctactgtccac
gaaatcctgtgcaagctcagcttggagggtgatcattctacacccccaagtgcctattga
```

Figure 33



BHK cells transfected with DEVD-caspase biosensor. (A) Cells before stimulation of apoptosis. (B) Another field of cells after stimulation with 250 μg/ml cis-platin (4 h).

FIGURE 34
SEQUENCE ID NO 3-4

Sequence and Translation: EYFP-DEVD-MAPKDM

This sequence codes for a caspase biosensor with EYFP (italic)as the fluorescent marker and the MAP4 projection domain retaining the chimera in the cytoplasm (unformatted text). These two regions are separate by a caspase-3 recognition site consisting of the sequence KGDEVDG(bold text).

```
+1   M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
                                        BanI
                                       -------
 1   ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC
     TACCACTCGTTC CCGCTCCTCGAC AAGTGGCCCCAC CACGGGTAGGAC CAGCTCGACCTG

+1   G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
61   GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC
     CCGCTGCATTTG CCGGTGTTCAAG TCGCACAGGCCG CTCCCGCTCCCG CTACGGTGGATG

+1   G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
121  GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC
     CCGTTCGACTGG GACTTCAAGTAG ACGTGGTGGCCG TTCGACGGGCAC GGGACCGGGTGG

+1   L  V  T  T  F  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K
181  CTCGTGACCACC TTCGGCTACGGC CTGCAGTGCTTC GCCCGCTACCCC GACCACATGAAG
     GAGCACTGGTGG AAGCCGATGCCG GACGTCACGAAG CGGGCGATGGGG CTGGTGTACTTC

+1   Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
241  CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC
     GTCGTGCTGAAG AAGTTCAGGCGG TACGGGCTTCCG ATGCAGGTCCTC GCGTGGTAGAAG

+1   F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
301  TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG
     AAGTTCCTGCTG CCGTTGATGTTC TGGGCGCGGCTC CACTTCAAGCTC CCGCTGTGGGAC

+1   V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
361  GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC
     CACTTGGCGTAG CTCGACTTCCCG TAGCTGAAGTTC CTCCTGCCGTTG TAGGACCCCGTG

+1   K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
421  AAGCTGGAGTAC AACTACAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC
     TTCGACCTCATG TTGATGTTGTCG GTGTTGCAGATA TAGTACCGGCTG TTCGTCTTCTTG
```

```
   +1  D  H  Y  Q   Q  N  T  P   I  G  D  G   P  V  L  L   P  D  N  H
541  GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC
     CTGGTGATGGTC GTCTTGTGGGGG TAGCCGCTGCCG GGGCACGACGAC GGGCTGTTGGTG

   +1  Y  L  S  Y   Q  S  A  L   S  K  D  P   N  E  K  R   D  H  M  V
601  TACCTGAGCTAC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC
     ATGGACTCGATG GTCAGGCGGGAC TCGTTTCTGGGG TTGCTCTTCGCG CTAGTGTACCAG

   +1  L  L  E  F   V  T  A  A   G  I  T  L   G  M  D  E   L  Y  K  K
661  CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAGAAG
     GACGACCTCAAG CACTGGCGGCGG CCCTAGTGAGAG CCGTACCTGCTC GACATGTTCTTC

   +1  G  D  E  V   D  G  A  D   L  S  L  V   D  A  L  T   E  P  P  P
                                                HincII
                                                ------
721  GGAGACGAAGTG GACGGAGCCGAC CTCAGTCTTGTG GATGCGTTGACA GAACCACCTCCA
     CCTCTGCTTCAC CTTCCTCGGCTG GAGTCAGAACAC CTACGCAACTGT CTTGGTGGAGGT

   +1  E  I  E  G   E  I  K  R   D  F  M  A   A  L  E  A   E  P  Y  D
781  GAAATTGAGGGA GAAATAAAGCGA GACTTCATGGCT GCGCTGGAGGCA GAGCCCTATGAT
     CTTTAACTCCCT CTTTATTTCGCT CTGAAGTACCGA CGCGACCTCCGT CTCGGGATACTA

   +1  D  I  V  G   E  T  V  E   K  T  E  F   I  P  L  L   D  G  D  E
841  GACATCGTGGGA GAAACTGTGGAG AAAACTGAGTTT ATTCCTCTCCTG GATGGTGATGAG
     CTGTAGCACCCT CTTTGACACCTC TTTTGACTCAAA TAAGGAGAGGAC CTACCACTACTC

   +1  K  T  G  N   S  E  S  K   K  K  P  C   L  D  T  S   Q  V  E  G
901  AAAACCGGGAAC TCAGAGTCCAAA AAGAAACCCTGC TTAGACACTAGC CAGGTTGAAGGT
     TTTTGGCCCTTG AGTCTCAGGTTT TTCTTTGGGACG AATCTGTGATCG GTCCAACTTCCA

   +1  I  P  S  S   K  P  T  L   L  A  N  G   D  H  G  M   E  G  N  N
961  ATCCCATCTTCT AAACCAACACTC CTAGCCAATGGT GATCATGGAATG GAGGGGAATAAC
     TAGGGTAGAAGA TTTGGTTGTGAG GATCGGTTACCA CTAGTACCTTAC CTCCCCTTATTG

   +1  T  A  G  S   P  T  D  F   L  E  E  R   V  D  Y  P   D  Y  Q  S
1021 ACTGCAGGGTCT CCAACTGACTTC CTTGAAGAGAGA GTGGACTATCCG GATTATCAGAGC
     TGACGTCCCAGA GGTTGACTGAAG GAACTTCTCTCT CACCTGATAGGC CTAATAGTCTCG

   +1  S  Q  N  W   P  E  D  A   S  F  C  F   Q  P  Q  Q   V  L  D  T
                                HindIII
                                --------
1081 AGCCAGAACTGG CCAGAAGATGCA AGCTTTTGTTTC CAGCCTCAGCAA GTGTTAGATACT
     TCGGTCTTGACC GGTCTTCTACGT TCGAAAACAAAG GTCGGAGTCGTT CACAATCTATGA

   +1  D  Q  A  E   P  F  N  E   H  R  D  D   G  L  A  D   L  L  F  V
                                                          BglII
                                                          --------
1141 GACCAGGCTGAG CCCTTTAACGAG CACCGTGATGAT GGTTTGGCAGAT CTGCTCTTTGTC
     CTGGTCCGACTC GGGAAATTGCTC GTGGCACTACTA CCAAACCGTCTA GACGAGAAACAG
```

```
   +1  Y  G  M  L   P  C  D  S   F  A  S  T   A  V  V  S   Q  E  W  S
1261  TACGGTATGCTT CCCTGTGACTCA TTTGCTTCCACG GCTGTTGTATCT CAGGAGTGGTCT
      ATGCCATACGAA GGGACACTGAGT AAACGAAGGTGC CGACAACATAGA GTCCTCACCAGA

   +1  V  G  A  P   N  S  P  C   S  E  S  C   V  S  P  E   V  T  I  E
1321  GTGGGAGCCCCA AACTCTCCATGT TCAGAGTCCTGT GTCTCCCCAGAG GTTACTATAGAA
      CACCCTCGGGGT TTGAGAGGTACA AGTCTCAGGACA CAGAGGGGTCTC CAATGATATCTT

   +1  T  L  Q  P   A  T  E  L   S  K  A  A   E  V  E  S   V  K  E  Q
1381  ACCCTACAGCCA GCAACAGAGCTC TCCAAGGCAGCA GAAGTGGAATCA GTGAAAGAGCAG
      TGGGATGTCGGT CGTTGTCTCGAG AGGTTCCGTCGT CTTCACCTTAGT CACTTTCTCGTC

   +1  L  P  A  K   A  L  E  T   M  A  E  Q   T  T  D  V   V  H  S  P
                                                   BstXI
                                                ~~~~~~~~~~~~~~
                                                                ApaLI
                                                               ~~~~~~
1441  CTGCCAGCTAAA GCATTGGAAACG ATGGCAGAGCAG ACCACTGATGTG GTGCACTCTCCA
      GACGGTCGATTT CGTAACCTTTGC TACCGTCTCGTC TGGTGACTACAC CACGTGAGAGGT

   +1  S  T  D  T   T  P  G  P   D  T  E  A   A  L  A  K   D  I  E  E
1501  TCCACAGACACA ACACCAGGCCCA GACACAGAGGCA GCACTGGCTAAA GACATAGAAGAG
      AGGTGTCTGTGT TGTGGTCCGGGT CTGTGTCTCCGT CGTGACCGATTT CTGTATCTTCTC

   +1  I  T  K  P   D  V  I  L   A  N  V  T   Q  P  S  T   E  S  D  M
1561  ATCACCAAGCCA GATGTGATATTG GCAAATGTCACG CAGCCATCTACT GAATCGGATATG
      TAGTGGTTCGGT CTACACTATAAC CGTTTACAGTGC GTCGGTAGATGA CTTAGCCTATAC

   +1  F  L  A  Q   D  M  E  L   L  T  G  T   E  A  A  H   A  N  N  I
1621  TTCCTGGCCCAG GACATGGAACTA CTCACAGGAACA GAGGCAGCCCAC GCTAACAATATC
      AAGGACCGGGTC CTGTACCTTGAT GAGTGTCCTTGT CTCCGTCGGGTG CGATTGTTATAG

   +1  I  L  P  T   E  P  D  E   S  S  T  K   D  V  A  P   P  M  E  E
1681  ATATTGCCTACA GAACCAGACGAA TCTTCAACCAAG GATGTAGCACCA CCTATGGAAGAA
      TATAACGGATGT CTTGGTCTGCTT AGAAGTTGGTTC CTACATCGTGGT GGATACCTTCTT

   +1  E  I  V  P   G  N  D  T   T  S  P  K   E  T  E  T   T  L  P  I
1741  GAAATTGTCCCA GGCAATGATACG ACATCCCCCAAA GAAACAGAGACA ACACTTCCAATA
      CTTTAACAGGGT CCGTTACTATGC TGTAGGGGGTTT CTTTGTCTCTGT TGTGAAGGTTAT

   +1  K  M  D  L   A  P  P  E   D  V  L  L   T  K  E  T   E  L  A  P
                  BanI
               ~~~~~~~
1801  AAAATGGACTTG GCACCACCTGAG GATGTGTTACTT ACCAAAGAAACA GAACTAGCCCCA
      TTTTACCTGAAC CGTGGTGGACTC CTACACAATGAA TGGTTTCTTTGT CTTGATCGGGGT

   +1  A  K  G  M   V  S  L  S   E  I  E  E   A  L  A  K   N  D  V  R
         BstXI
      ~~~~~~~~~~~~~~
```

```
      AGACGTCTTTAT GGACACCGAGTC CTCTGTCACCAG AGTCTTTGTCTC CACCAGGACCGT

   +1 T  E  V  V  L  P  S  D  P  I  T  T  L  T  K  D  V  T  L  P
1981 ACAGAAGTGGTA CTGCCCTCAGAT CCCATAACAACA TTGACAAAGGAT GTGACACTCCCC
     TGTCTTCACCAT GACGGGAGTCTA GGGTATTGTTGT AACTGTTTCCTA CACTGTGAGGGG

   +1 L  E  A  E  R  P  L  V  T  D  M  T  P  S  L  E  T  E  M  T
2041 TTAGAAGCAGAG AGACCGTTGGTG ACGGACATGACT CCATCTCTGGAA ACAGAAATGACC
     AATCTTCGTCTC TCTGGCAACCAC TGCCTGTACTGA GGTAGAGACCTT TGTCTTTACTGG

   +1 L  G  K  E  T  A  P  P  T  E  T  N  L  G  M  A  K  D  M  S
                                        ApoI
                                        -------
2101 CTAGGCAAAGAG ACAGCTCCACCC ACAGAAACAAAT TTGGGCATGGCC AAAGACATGTCT
     GATCCGTTTCTC TGTCGAGGTGGG TGTCTTTGTTTA AACCCGTACCGG TTTCTGTACAGA

   +1 P  L  P  E  S  E  V  T  L  G  K  D  V  V  I  L  P  E  T  K
2161 CCACTCCCAGAA TCAGAAGTGACT CTGGGCAAGGAC GTGGTTATACTT CCAGAAACAAAG
     GGTGAGGGTCTT AGTCTTCACTGA GACCCGTTCCTG CACCAATATGAA GGTCTTTGTTTC

   +1 V  A  E  F  N  N  V  T  P  L  S  E  E  E  V  T  S  V  K  D
2221 GTGGCTGAGTTT AACAATGTGACT CCACTTTCAGAA GAAGAGGTAACC TCAGTCAAGGAC
     CACCGACTCAAA TTGTTACACTGA GGTGAAAGTCTT CTTCTCCATTGG AGTCAGTTCCTG

   +1 M  S  P  S  A  E  T  E  A  P  L  A  K  N  A  D  L  H  S  G
2281 ATGTCTCCGTCT GCAGAAACAGAG GCTCCCCTGGCT AAGAATGCTGAT CTGCACTCAGGA
     TACAGAGGCAGA CGTCTTTGTCTC CGAGGGGACCGA TTCTTACGACTA GACGTGAGTCCT

   +1 T  E  L  I  V  D  N  S  M  A  P  A  S  D  L  A  L  P  L  E
2341 ACAGAGCTGATT GTGGACAACAGC ATGGCTCCAGCC TCCGATCTTGCA CTGCCCTTGGAA
     TGTCTCGACTAA CACCTGTTGTCG TACCGAGGTCGG AGGCTAGAACGT GACGGGAACCTT

   +1 T  K  V  A  T  V  P  I  K  D  K  G  *
2401 ACAAAAGTAGCA ACAGTTCCAATT AAAGACAAAGGA TGA
     TGTTTTCATCGT TGTCAAGGTTAA TTTCTGTTTCCT ACT
```

FIGURE 35
SEQUENCE ID NO 5-6

Sequence and Translation: EYFP-DEAD-MAPKDM

This sequence codes for a caspase biosensor with EYFP (italic)as the fluorescent marker and the MAP4 projection domain retaining the chimera in the cytoplasm (unformatted text). These two regions are separate by a caspase-3 recognition site consisting of the sequence PRDEADS(bold text).

```
+1  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
                                        BanI
                                       -------
 1  ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC
    TACCACTCGTTC CCGCTCCTCGAC AAGTGGCCCCAC CACGGGTAGGAC CAGCTCGACCTG

+1  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
61  GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC
    CCGCTGCATTTG CCGGTGTTCAAG TCGCACAGGCCG CTCCCGCTCCCG CTACGGTGGATG

+1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
121 GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC
    CCGTTCGACTGG GACTTCAAGTAG ACGTGGTGGCCG TTCGACGGGCAC GGGACCGGGTGG

+1  L  V  T  T  F  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K
181 CTCGTGACCACC TTCGGCTACGGC CTGCAGTGCTTC GCCCGCTACCCC GACCACATGAAG
    GAGCACTGGTGG AAGCCGATGCCG GACGTCACGAAG CGGGCGATGGGG CTGGTGTACTTC

+1  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
241 CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC
    GTCGTGCTGAAG AAGTTCAGGCGG TACGGGCTTCCG ATGCAGGTCCTC GCGTGGTAGAAG

+1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
301 TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG
    AAGTTCCTGCTG CCGTTGATGTTC TGGGCGCGGCTC CACTTCAAGCTC CCGCTGTGGGAC

+1  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
361 GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC
    CACTTGGCGTAG CTCGACTTCCCG TAGCTGAAGTTC CTCCTGCCGTTG TAGGACCCCGTG

+1  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
421 AAGCTGGAGTAC AACTACAACAGC CACAACGTCTAT ATCATGGCCGAC AAGCAGAAGAAC
    TTCGACCTCATG TTGATGTTGTCG GTGTTGCAGATA TAGTACCGGCTG TTCGTCTTCTTG
```

```
  +1  D  H  Y  Q   Q  N  T  P   I  G  D  G   P  V  L  L   P  D  N  H
541  GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC
     CTGGTGATGGTC GTCTTGTGGGGG TAGCCGCTGCCG GGGCACGACGAC GGGCTGTTGGTG

  +1  Y  L  S  Y   Q  S  A  L   S  K  D  P   N  E  K  R   D  H  M  V
601  TACCTGAGCTAC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC
     ATGGACTCGATG GTCAGGCGGGAC TCGTTTCTGGGG TTGCTCTTCGCG CTAGTGTACCAG

  +1  L  L  E  F   V  T  A  A   G  I  T  L   G  M  D  E   L  Y  K  P
661  CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAGCCC
     GACGACCTCAAG CACTGGCGGCGG CCCTAGTGAGAG CCGTACCTGCTC GACATGTTCGGG

  +1  R  D  E  A   D  S  A  D   L  S  L  V   D  A  L  T   E  P  P  P
                                                HincII
                                                ------
721  AGAGACGAAGCC GACAGCGCCGAC CTCAGTCTTGTG GATGCGTTGACA GAACCACCTCCA
     TCTCTGCTTCGG CTGTCGCGGCTG GAGTCAGAACAC CTACGCAACTGT CTTGGTGGAGGT

  +1  E  I  E  G   E  I  K  R   D  F  M  A   A  L  E  A   E  P  Y  D
781  GAAATTGAGGGA GAAATAAAGCGA GACTTCATGGCT GCGCTGGAGGCA GAGCCCTATGAT
     CTTTAACTCCCT CTTTATTTCGCT CTGAAGTACCGA CGCGACCTCCGT CTCGGGATACTA

  +1  D  I  V  G   E  T  V  E   K  T  E  F   I  P  L  L   D  G  D  E
841  GACATCGTGGGA GAAACTGTGGAG AAAACTGAGTTT ATTCCTCTCCTG GATGGTGATGAG
     CTGTAGCACCCT CTTTGACACCTC TTTTGACTCAAA TAAGGAGAGGAC CTACCACTACTC

  +1  K  T  G  N   S  E  S  K   K  K  P  C   L  D  T  S   Q  V  E  G
901  AAAACCGGGAAC TCAGAGTCCAAA AAGAAACCCTGC TTAGACACTAGC CAGGTTGAAGGT
     TTTTGGCCCTTG AGTCTCAGGTTT TTCTTTGGGACG AATCTGTGATCG GTCCAACTTCCA

  +1  I  P  S  S   K  P  T  L   L  A  N  G   D  H  G  M   E  G  N  N
961  ATCCCATCTTCT AAACCAACACTC CTAGCCAATGGT GATCATGGAATG GAGGGGAATAAC
     TAGGGTAGAAGA TTTGGTTGTGAG GATCGGTTACCA CTAGTACCTTAC CTCCCCTTATTG

  +1  T  A  G  S   P  T  D  F   L  E  E  R   V  D  Y  P   D  Y  Q  S
1021 ACTGCAGGGTCT CCAACTGACTTC CTTGAAGAGAGA GTGGACTATCCG GATTATCAGAGC
     TGACGTCCCAGA GGTTGACTGAAG GAACTTCTCTCT CACCTGATAGGC CTAATAGTCTCG

  +1  S  Q  N  W   P  E  D  A   S  F  C  F   Q  P  Q  Q   V  L  D  T
                               HindIII
                               -------
1081 AGCCAGAACTGG CCAGAAGATGCA AGCTTTTGTTTC CAGCCTCAGCAA GTGTTAGATACT
     TCGGTCTTGACC GGTCTTCTACGT TCGAAAACAAAG GTCGGAGTCGTT CACAATCTATGA

  +1  D  Q  A  E   P  F  N  E   H  R  D  D   G  L  A  D   L  L  F  V
                                                        BglII
                                                        -------
1141 GACCAGGCTGAG CCCTTTAACGAG CACCGTGATGAT GGTTTGGCAGAT CTGCTCTTTGTC
```

```
   +1  Y  G  M  L   P  C  D  S   F  A  S  T   A  V  V  S   Q  E  W  S
1261  TACGGTATGCTT CCCTGTGACTCA TTTGCTTCCACG GCTGTTGTATCT CAGGAGTGGTCT
      ATGCCATACGAA GGGACACTGAGT AAACGAAGGTGC CGACAACATAGA GTCCTCACCAGA

   +1  V  G  A  P   N  S  P  C   S  E  S  C   V  S  P  E   V  T  I  E
1321  GTGGGAGCCCCA AACTCTCCATGT TCAGAGTCCTGT GTCTCCCCAGAG GTTACTATAGAA
      CACCCTCGGGGT TTGAGAGGTACA AGTCTCAGGACA CAGAGGGGTCTC CAATGATATCTT

   +1  T  L  Q  P   A  T  E  L   S  K  A  A   E  V  E  S   V  K  E  Q
1381  ACCCTACAGCCA GCAACAGAGCTC TCCAAGGCAGCA GAAGTGGAATCA GTGAAAGAGCAG
      TGGGATGTCGGT CGTTGTCTCGAG AGGTTCCGTCGT CTTCACCTTAGT CACTTTCTCGTC

   +1  L  P  A  K   A  L  E  T   M  A  E  Q   T  T  D  V   V  H  S  P
                                                  BstXI
                                               -------------
                                                           ApaLI
                                                           ------
1441  CTGCCAGCTAAA GCATTGGAAACG ATGGCAGAGCAG ACCACTGATGTG GTGCACTCTCCA
      GACGGTCGATTT CGTAACCTTTGC TACCGTCTCGTC TGGTGACTACAC CACGTGAGAGGT

   +1  S  T  D  T   T  P  G  P   D  T  E  A   A  L  A  K   D  I  E  E
1501  TCCACAGACACA ACACCAGGCCCA GACACAGAGGCA GCACTGGCTAAA GACATAGAAGAG
      AGGTGTCTGTGT TGTGGTCCGGGT CTGTGTCTCCGT CGTGACCGATTT CTGTATCTTCTC

   +1  I  T  K  P   D  V  I  L   A  N  V  T   Q  P  S  T   E  S  D  M
1561  ATCACCAAGCCA GATGTGATATTG GCAAATGTCACG CAGCCATCTACT GAATCGGATATG
      TAGTGGTTCGGT CTACACTATAAC CGTTTACAGTGC GTCGGTAGATGA CTTAGCCTATAC

   +1  F  L  A  Q   D  M  E  L   L  T  G  T   E  A  A  H   A  N  N  I
1621  TTCCTGGCCCAG GACATGGAACTA CTCACAGGAACA GAGGCAGCCCAC GCTAACAATATC
      AAGGACCGGGTC CTGTACCTTGAT GAGTGTCCTTGT CTCCGTCGGGTG CGATTGTTATAG

   +1  I  L  P  T   E  P  D  E   S  S  T  K   D  V  A  P   P  M  E  E
1681  ATATTGCCTACA GAACCAGACGAA TCTTCAACCAAG GATGTAGCACCA CCTATGGAAGAA
      TATAACGGATGT CTTGGTCTGCTT AGAAGTTGGTTC CTACATCGTGGT GGATACCTTCTT

   +1  E  I  V  P   G  N  D  T   T  S  P  K   E  T  E  T   T  L  P  I
1741  GAAATTGTCCCA GGCAATGATACG ACATCCCCCAAA GAAACAGAGACA ACACTTCCAATA
      CTTTAACAGGGT CCGTTACTATGC TGTAGGGGGTTT CTTTGTCTCTGT TGTGAAGGTTAT

   +1  K  M  D  L   A  P  P  E   D  V  L  L   T  K  E  T   E  L  A  P
                  BanI
               --------
1801  AAAATGGACTTG GCACCACCTGAG GATGTGTTACTT ACCAAAGAAACA GAACTAGCCCCA
      TTTTACCTGAAC CGTGGTGGACTC CTACACAATGAA TGGTTTCTTTGT CTTGATCGGGGT

   +1  A  K  G  M   V  S  L  S   E  I  E  E   A  L  A  K   N  D  V  R
         BstXI
```

```
1921  TCTGCAGAAATA CCTGTGGCTCAG GAGACAGTGGTC TCAGAAACAGAG GTGGTCCTGGCA
      AGACGTCTTTAT GGACACCGAGTC CTCTGTCACCAG AGTCTTTGTCTC CACCAGGACCGT

  +1  T  E  V  V   L  P  S  D   P  I  T  T   L  T  K  D   V  T  L  P
1981  ACAGAAGTGGTA CTGCCCTCAGAT CCCATAACAACA TTGACAAAGGAT GTGACACTCCCC
      TGTCTTCACCAT GACGGGAGTCTA GGGTATTGTTGT AACTGTTTCCTA CACTGTGAGGGG

  +1  L  E  A  E   R  P  L  V   T  D  M  T   P  S  L  E   T  E  M  T
2041  TTAGAAGCAGAG AGACCGTTGGTG ACGGACATGACT CCATCTCTGGAA ACAGAAATGACC
      AATCTTCGTCTC TCTGGCAACCAC TGCCTGTACTGA GGTAGAGACCTT TGTCTTTACTGG

  +1  L  G  K  E   T  A  P  P   T  E  T  N   L  G  M  A   K  D  M  S
                                          ApoI
                                          -------
2101  CTAGGCAAAGAG ACAGCTCCACCC ACAGAAACAAAT TTGGGCATGGCC AAAGACATGTCT
      GATCCGTTTCTC TGTCGAGGTGGG TGTCTTTGTTTA AACCCGTACCGG TTTCTGTACAGA

  +1  P  L  P  E   S  E  V  T   L  G  K  D   V  V  I  L   P  E  T  K
2161  CCACTCCCAGAA TCAGAAGTGACT CTGGGCAAGGAC GTGGTTATACTT CCAGAAACAAAG
      GGTGAGGGTCTT AGTCTTCACTGA GACCCGTTCCTG CACCAATATGAA GGTCTTTGTTTC

  +1  V  A  E  F   N  N  V  T   P  L  S  E   E  E  V  T   S  V  K  D
2221  GTGGCTGAGTTT AACAATGTGACT CCACTTTCAGAA GAAGAGGTAACC TCAGTCAAGGAC
      CACCGACTCAAA TTGTTACACTGA GGTGAAAGTCTT CTTCTCCATTGG AGTCAGTTCCTG

  +1  M  S  P  S   A  E  T  E   A  P  L  A   K  N  A  D   L  H  S  G
2281  ATGTCTCCGTCT GCAGAAACAGAG GCTCCCCTGGCT AAGAATGCTGAT CTGCACTCAGGA
      TACAGAGGCAGA CGTCTTTGTCTC CGAGGGGACCGA TTCTTACGACTA GACGTGAGTCCT

  +1  T  E  L  I   V  D  N  S   M  A  P  A   S  D  L  A   L  P  L  E
2341  ACAGAGCTGATT GTGGACAACAGC ATGGCTCCAGCC TCCGATCTTGCA CTGCCCTTGGAA
      TGTCTCGACTAA CACCTGTTGTCG TACCGAGGTCGG AGGCTAGAACGT GACGGGAACCTT

  +1  T  K  V  A   T  V  P  I   K  D  K  G   *
2401  ACAAAAGTAGCA ACAGTTCCAATT AAAGACAAAGGA TGA
      TGTTTTCATCGT TGTCAAGGTTAA TTTCTGTTTCCT ACT
```

FIGURE 36
SEQUENCE ID NO 7-8

Sequence and Translation: F25-MEK1

This sequence codes for a chimeric molecule that reports activity of a particular zinc metalloprotease. The molecule consists of GFP (underline) and human MEK1 cDNA (double underline). The cleavage site is shown in bold and the nuclear export sequence (NES) is shown in italic.

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
                NheI
                -------
    1  ATGGCTAGCAAA GGAGAAGAACTC TTCACTGGAGTT GTCCCAATTCTT GTTGAATTAGAT
       TACCGATCGTTT CCTCTTCTTGAG AAGTGACCTCAA CAGGGTTAAGAA CAACTTAATCTA

      +1  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
   61  GGTGATGTTAAC GGCCACAAGTTC TCTGTCAGTGGA GAGGGTGAAGGT GATGCAACATAC
       CCACTACAATTG CCGGTGTTCAAG AGACAGTCACCT CTCCCACTTCCA CTACGTTGTATG

      +1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
  121  GGAAAACTTACC CTGAAGTTCATC TGCACTACTGGC AAACTGCCTGTT CCATGGCCAACA
       CCTTTTGAATGG GACTTCAAGTAG ACGTGATGACCG TTTGACGGACAA GGTACCGGTTGT

      +1  L  V  T  T  L  C  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K
                                                              NdeI
                                                              -------
  181  CTAGTCACTACT CTGTGCTATGGT GTTCAATGCTTT TCAAGATACCCG GATCATATGAAA
       GATCAGTGATGA GACACGATACCA CAAGTTACGAAA AGTTCTATGGGC CTAGTATACTTT

      +1  R  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
  241  CGGCATGACTTT TTCAAGAGTGCC ATGCCCGAAGGT TATGTACAGGAA AGGACCATCTTC
       GCCGTACTGAAA AAGTTCTCACGG TACGGGCTTCCA ATACATGTCCTT TCCTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
  301  TTCAAAGATGAC GGCAACTACAAG ACACGTGCTGAA GTCAAGTTTGAA GGTGATACCCTT
       AAGTTTCTACTG CCGTTGATGTTC TGTGCACGACTT CAGTTCAAACTT CCACTATGGGAA

      +1  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
  361  GTTAATAGAATC GAGTTAAAAGGT ATTGACTTCAAG GAAGATGGCAAC ATTCTGGGACAC
       CAATTATCTTAG CTCAATTTTCCA TAACTGAAGTTC CTTCTACCGTTG TAAGACCCTGTG

      +1  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
                                   AccI
                                   -------
  421  AAATTGGAATAC AACTATAACTCA CACAATGTATAC ATCATGGCAGAC AAACAAAAGAAT
```

```
        CCTTAGTTTCAC TTGAAGTTCTGG GCGGTGTTGTAA CTTCTACCTTCG CAAGTTGATCGT

    +1  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
541 GACCATTATCAA CAAAATACTCCA ATTGGCGATGGC CCTGTCCTTTTA CCAGACAACCAT
    CTGGTAATAGTT GTTTTATGAGGT TAACCGCTACCG GGACAGGAAAAT GGTCTGTTGGTA

    +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
                                    BstYI
                                    ------
601 TACCTGTCCACA CAATCTGCCCTT TCGAAAGATCCC AACGAAAAGAGA GACCACATGGTC
    ATGGACAGGTGT GTTAGACGGGAA AGCTTTCTAGGG TTGCTTTTCTCT CTGGTGTACCAG

    +1  L  L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  N  T
                                                                AgeI
                                                                ---
661 CTTCTTGAGTTT GTAACAGCTGCT GGGATTACACAT GGCATGGATGAA CTGTACAACACC
    GAAGAACTCAAA CATTGTCGACGA CCCTAATGTGTA CCGTACCTACTT GACATGTTGTGG

    +1  G  M  P  K  K  K  P  T  P  I  Q  L  N  P  A  P  D  G  S  A
        AgeI                                                    PstI
        ---                                                     -----
721 GGTATGCCCAAG AAGAAGCCGACG CCCATCCAGCTG AACCCGGCCCCC GACGGCTCTGCA
    CCATACGGGTTC TTCTTCGGCTGC GGGTAGGTCGAC TTGGGCCGGGGG CTGCCGAGACGT

    +1  V  N  G  T  S  S  A  E  T  N  L  E  A  L  Q  K  K  L  E  E
        PstI
        -
781 GTTAACGGGACC AGCTCTGCGGAG ACCAACTTGGAG GCCTTGCAGAAG AAGCTGGAGGAG
    CAATTGCCCTGG TCGAGACGCCTC TGGTTGAACCTC CGGAACGTCTTC TTCGACCTCCTC

    +1  L  E  L  D  E  Q  Q
841 CTAGAGCTTGAT GAGCAGCAGTGA
```

FIGURE 37
SEQUENCE ID NO 9-10

Sequence and Translation: F25-MEK2

This sequence codes for a chimeric molecule that reports activity of a particular zinc metalloprotease. The molecule consists of GFP (underline) and human MEK2 cDNA (double underline). The cleavage site is shown in bold and the nuclear export sequence (NES) is shown in italic.

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
                NheI
                -------
       1  ATGGCTAGCAAA GGAGAAGAACTC TTCACTGGAGTT GTCCCAATTCTT GTTGAATTAGAT
          TACCGATCGTTT CCTCTTCTTGAG AAGTGACCTCAA CAGGGTTAAGAA CAACTTAATCTA

      +1  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
                   HincII
                   -------
      61  GGTGATGTTAAC GGCCACAAGTTC TCTGTCAGTGGA GAGGGTGAAGGT GATGCAACATAC
          CCACTACAATTG CCGGTGTTCAAG AGACAGTCACCT CTCCCACTTCCA CTACGTTGTATG

      +1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
     121  GGAAAACTTACC CTGAAGTTCATC TGCACTACTGGC AAACTGCCTGTT CCATGGCCAACA
          CCTTTTGAATGG GACTTCAAGTAG ACGTGATGACCG TTTGACGGACAA GGTACCGGTTGT

      +1  L  V  T  T  L  C  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K
                                                            NdeI
                                                            -------
     181  CTAGTCACTACT CTGTGCTATGGT GTTCAATGCTTT TCAAGATACCCG GATCATATGAAA
          GATCAGTGATGA GACACGATACCA CAAGTTACGAAA AGTTCTATGGGC CTAGTATACTTT

      +1  R  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     241  CGGCATGACTTT TTCAAGAGTGCC ATGCCCGAAGGT TATGTACAGGAA AGGACCATCTTC
          GCCGTACTGAAA AAGTTCTCACGG TACGGGCTTCCA ATACATGTCCTT TCCTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
     301  TTCAAAGATGAC GGCAACTACAAG ACACGTGCTGAA GTCAAGTTTGAA GGTGATACCCTT
          AAGTTTCTACTG CCGTTGATGTTC TGTGCACGACTT CAGTTCAAACTT CCACTATGGGAA

      +1  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
     361  GTTAATAGAATC GAGTTAAAAGGT ATTGACTTCAAG GAAGATGGCAAC ATTCTGGGACAC
          CAATTATCTTAG CTCAATTTTCCA TAACTGAAGTTC CTTCTACCGTTG TAAGACCCTGTG

      +1  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N
```

```
      +1  G  I  K  V  N  F  K  T  R  H  N  I  E  D  G  S  V  Q  L  A
481   GGAATCAAAGTG AACTTCAAGACC CGCCACAACATT GAAGATGGAAGC GTTCAACTAGCA
      CCTTAGTTTCAC TTGAAGTTCTGG GCGGTGTTGTAA CTTCTACCTTCG CAAGTTGATCGT

      +1  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
541   GACCATTATCAA CAAAATACTCCA ATTGGCGATGGC CCTGTCCTTTTA CCAGACAACCAT
      CTGGTAATAGTT GTTTTATGAGGT TAACCGCTACCG GGACAGGAAAAT GGTCTGTTGGTA

      +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V
601   TACCTGTCCACA CAATCTGCCCTT TCGAAAGATCCC AACGAAAAGAGA GACCACATGGTC
      ATGGACAGGTGT GTTAGACGGGAA AGCTTTCTAGGG TTGCTTTTCTCT CTGGTGTACCAG

      +1  L  L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  N  T
                                                                AgeI
                                                                ---
661   CTTCTTGAGTTT GTAACAGCTGCT GGGATTACACAT GGCATGGATGAA CTGTACAACACC
      GAAGAACTCAAA CATTGTCGACGA CCCTAATGTGTA CCGTACCTACTT GACATGTTGTGG

      +1  G  M  L  A  R  R  K  P  V  L  P  A  L  T  I  N  P  T  I  A
          AgeI
          ---
721   GGTATGCTGGCC CGGAGGAAGCCG GTGCTGCCGGCG CTCACCATCAAC CCTACCATCGCC
      CCATACGACCGG GCCTCCTTCGGC CACGACGGCCGC GAGTGGTAGTTG GGATGGTAGCGG

      +1  E  G  P  S  P  T  S  E  G  A  S  E  A  N  L  V  D  L  Q  K
              BanII
              -------
              ApaI                          BglI                  PstI
              ------                        ----------            ------
781   GAGGGCCCATCC CCTACCAGCGAG GGCGCCTCCGAG GCAAACCTGGTG GACCTGCAGAAG
      CTCCCGGGTAGG GGATGGTCGCTC CCGCGGAGGCTC CGTTTGGACCAC CTGGACGTCTTC

      +1  K  L  E  E  L  E  L  D  E  Q  Q
841   AAGCTGGAGGAG CTGGAACTTGAC GAGCAGCAGTAA
```

Figure 38

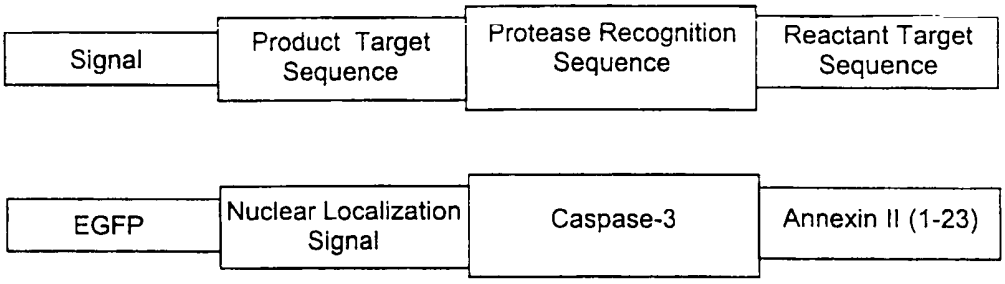

Signal
Product Target Sequence
Protease Recognition Sequence
Reactant Target Sequence
EGFP
Nuclear Localization Signal
Caspase-3
Annexin II (1-23)

FIGURE 39
SEQUENCE ID NO 11-12
Caspase 3 – DEVD- substrate

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1  ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
          TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51  TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
          ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101  GTCAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
          CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                              NcoI
                             --------
     151  ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
          TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
     201  TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
          ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251  TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
          AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301  TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
          AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351  TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
          ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401  ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
          TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451  GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA
          CATATGTAGT ACCGTCTGTT TGTTTTCTTA CCTTAGTTTC ACTTGAAGTT

      +1  T  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y
     501  GACCCGCCAC AACATTGAAG ATGGAAGCGT TCAACTAGCA GACCATTATC
          CTGGGCGGTG TTGTAACTTC TACCTTCGCA AGTTGATCGT CTGGTAATAG
```

```
      +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
601   TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
      ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

      +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
651   CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
      GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

      +1 M  D  E  L  Y  N  S  G  R  R  K  R  Q  K  R  S  A
701   TGGATGAACT GTACAACTCC GGAAGAAGGA AACGACAAAA GCGATCGGCT
      ACCTACTTGA CATGTTGAGG CCTTCTTCCT TTGCTGTTTT CGCTAGCCGA

      +1  V  K  S  E  G  K  R  K  C  D  E  V  D  G  I  D  E
751   GTTAAATCTG AAGGAAAGAG AAAGTGTGAC GAAGTTGATG GAATTGATGA
      CAATTTAGAC TTCCTTTCTC TTTCACACTG CTTCAACTAC CTTAACTACT

      +1  V  A  S  T  M  S  T  V  H  E  I  L  C  K  L  S
801   AGTAGCAAGT ACTATGTCTA CTGTCCACGA AATCCTGTGC AAGCTCAGCT
      TCATCGTTCA TGATACAGAT GACAGGTGCT TTAGGACACG TTCGAGTCGA

      +1 L  E  G  V  H  S  T  P  P  S  T  R  I
                                              BamHI
                                              -----
851   TGGAGGGTGT TCATTCTACA CCCCCAAGTA CCCGGATCC
      ACCTCCCACA AGTAAGATGT GGGGGTTCAT GGGCCTAGG
```

FIGURE 40

Seq. ID. NO. 13-14

Caspase 6 – VEID- sub.

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1  ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
          TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51  TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
          ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101  GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
          CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                              NcoI
                              --------
     151  ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
          TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
     201  TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
          ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251  TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
          AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301  TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
          AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351  TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
          ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401  ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
          TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451  GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA
          CATATGTAGT ACCGTCTGTT TGTTTTCTTA CCTTAGTTTC ACTTGAAGTT

      +1  T  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y
     501  GACGGCGCAC AACATTGAAG ATGGAAGCGT TCAACTAGCA GACCATTATC
```

```
 551  AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
      TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA

       +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
 601  TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
      ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

       +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
 651  CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
      GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

       +1  M  D  E  L  Y  N  S  G  R  R  K  R  Q  K  R  S  T
 701  TGGATGAACT GTACAACTCC GGAAGAAGGA AACGACAAAA GCGATCGACA
      ACCTACTTGA CATGTTGAGG CCTTCTTCCT TTGCTGTTTT CGCTAGCTGT

       +1  R  L  V  E  I  D  N  S  T  M  S  T  V  H  E  I  L
 751  AGACTTGTTG AAATTGACAA CAGTACTATG AGCACAGTAC ACGAAATTTT
      TCTGAACAAC TTTAACTGTT GTCATGATAC TCGTGTCATG TGCTTTAAAA

       +1  C  K  L  S  L  E  G  V  H  S  T  P  P  S  A  G  T
                  HindIII
                  -------
 801  ATGTAAATTA AGCTTAGAAG GAGTACACAG TACACCACCA AGCGCA
      TACATTTAAT TCGAATCTTC CTCATGTGTC ATGTGGTGGT TCGCGT
```

FIGURE 41

SEQ ID NO:15-16

Caspase 8 - VETD

```
      +1 M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1 ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
         TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51 TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
         ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1 G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101 GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
         CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1 T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                          NcoI
                          -------
     151 ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
         TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
     201 TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
         ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1 F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251 TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
         AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1 F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301 TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
         AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351 TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
         ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1 D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401 ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
         TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1 V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451 GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA
         CATATGTAGT ACCGTCTGTT TGTTTTCTTA CCTTAGTTTC ACTTGAAGTT
```

```
      +1 Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
  551 AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
      TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA

      +1 Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
  601 TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
      ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

      +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
  651 CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
      GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

      +1 M  D  E  L  Y  N  S  G  R  S  K  R  Q  K  R  S
  701 TGGATGAACT GTACAAC TCCGGAAGAA GCAAACGACA AAAGCGATCG
      ACCTACTTGA CATGTTG AGGCCTTCTT CGTTTGCTGT TTTCGCTAGC

      +1 Y  E  K  G  I  P  V  E  T  D  S  E  E  Q  A  Y  S
                                                HindIII
                                                -------
  751 TATGAAAAAG GAATACCAGT TGAAACAGAC AGCGAAGAGC AAGCTTATAG
      ATACTTTTTC CTTATGGTCA ACTTTGTCTG TCGCTTCTCG TTCGAATATC

      +1  T  M  S  T  V  H  E  I  L  C  K  L  S  L  E  G
  801 TACTATGTCT ACTGTCCACG AAATCCTGTG CAAGCTCAGC TTGGAGGGTG
      ATGATACAGA TGACAGGTGC TTTAGGACAC GTTCGAGTCG AACCTCCCAC

      +1 V  H  S  T  P  P  S  A  G  S
                                BamHI
                                ------
  851 TTCATTCTAC ACCCCCAAGT GCCGGATCC
      AAGTAAGATG TGGGGGTTCA CGGCCTAGG
```

FIGURE 42

SEQ ID NO:17-18

Sequence and Translation of Cas3- multiple DEVD

```
        +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
         1  ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
            TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA
        -1  H  S  A  F  S  F  F  E  E  S  S  N  D  W  N  K  N

        +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
        51  TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
            ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC
        -1  F  *  I  T  I  N  V  A  V  L  E  R  D  T  S  L

        +1  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
       101  GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
            CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA
        -1  T  F  T  I  C  C  V  S  F  K  G  Q  L  E  D  A  S

        +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                                NcoI
                               --------
       151  ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
            TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT
        -1  S  A  F  Q  R  N  W  P  W  C  *  D  S  S  Q  A  I

        +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
       201  TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
            ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA
        -1  T  N  L  A  K  *  S  V  R  I  M  H  F  P  M  V

        +1  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
       251  TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
            AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG
        -1  K  E  L  T  G  H  G  F  T  I  Y  L  F  P  G  D  E

        +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
       301  TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
            AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC
        -1  E  F  I  V  A  V  V  L  C  T  S  F  D  L  K  F  T

        +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
       351  TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
            ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC
        -1  I  G  K  N  I  S  D  L  *  F  T  N  V  E  L  F
```

```
      +1 V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451 GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA
         CATATGTAGT ACCGTCTGTT TGTTTTCTTA CCTTAGTTTC ACTTGAAGTT
      -1 Y  V  D  H  C  V  F  L  L  I  S  D  F  H  V  E  L

      +1  T  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y
     501 GACCCGCCAC AACATTGAAG ATGGAAGCGT TCAACTAGCA GACCATTATC
         CTGGGCGGTG TTGTAACTTC TACCTTCGCA AGTTGATCGT CTGGTAATAG
      -1  G  A  V  V  N  F  I  S  A  N  L  *  C  V  M  I

      +1 Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
     551 AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
         TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA
      -1 L  L  I  S  W  N  A  I  A  R  D  K  *  W  V  V  M

      +1 Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
     601 TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
         ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT
      -1 V  Q  G  C  L  R  G  K  R  F  I  G  V  F  L  S  V

      +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
     651 CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
         GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT
      -1  V  H  D  K  K  L  K  Y  C  S  S  P  N  C  M  A

      +1 M  D  E  L  Y  N     S  G  R  R  K  R  Q  K  R  S
     701 TGGATGAACT GTACAAC TCCGGAAGAA GGAAACGACA AAAGCGATCG
         ACCTACTTGA CATGTTG AGGCCTTCTT CCTTTGCTGT TTTCGCTAGC
      -1 H  I  F  Q  V  V  L  G  S  S  P  F  S  L  L  S  R

      +1 A  G  D  E  V  D  A  G  D  E  V  D  A  G  D  E  V
     751 GCAGGTGACG AAGTTGATGC AGGTGACGAA GTTGATGCAG GTGACGAAGT
         CGTCCACTGC TTCAACTACG TCCACTGCTT CAACTACGTC CACTGCTTCA
      -1 C  T  V  F  N  I  C  T  V  F  N  I  C  T  V  F  N

      +1  D  A  G  D  E  V  D  A  G  S  T  M  S  T  V  H
     801 TGATGCAGGT GACGAAGTTG ACGCAGGTAG TACTATGTCT ACTGTCCACG
         ACTACGTCCA CTGCTTCAAC TGCGTCCATC ATGATACAGA TGACAGGTGC
      -1  I  C  T  V  F  N  V  C  T  T  S  H  R  S  D  V

      +1 E  I  L  C  K  L  S  L  E  G  V  H  S  T  P  P  S
     851 AAATCCTGTG CAAGCTCAGC TTGGAGGGTG TTCATTCTAC ACCCCCAAGT
         TTTAGGACAC GTTCGAGTCG AACCTCCCAC AAGTAAGATG TGGGGGTTCA
      -1 F  D  Q  A  L  E  A  Q  L  T  N  M  R  C  G  W  T

      +1 A  G  S
            BamHI
           ------
```

FIGURE 43

SEQ ID 19-20

Caspase 8 –multiple vetd

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1  ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
          TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51  TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
          ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1 G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101  GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
          CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                              NcoI
                              --------
     151  ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
          TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
     201  TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
          ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1 F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251  TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
          AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301  TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
          AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351  TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
          ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1 D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401  ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
          TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451  GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA
          CATATGTAGT ACCGTCTGTT TGTTTTCTTA CCTTAGTTTC ACTTGAAGTT
```

```
      +1 Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
  551 AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
      TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA

      +1 Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
  601 TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
      ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

      +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
  651 CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
      GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

      +1 M  D  E  L  Y  N  S  G  R  R  K  R  Q  K  R  S
  701 TGGATGAACT GTACAAC TCCGGAAGAA GGAAACGACA AAAGCGATCG
      ACCTACTTGA CATGTTG AGGCCTTCTT CCTTTGCTGT TTTCGCTAGC

      +1 A  G  V  E  T  D  A  G  V  E  T  D  A  G  V  E  T
  751 GCAGGTGTTG AAACAGACGC AGGTGTTGAA ACAGACGCAG GTGTTGAAAC
      CGTCCACAAC TTTGTCTGCG TCCACAACTT TGTCTGCGTC CACAACTTTG

      +1  D  A  G  V  E  T  D  A  G  S  T  M  S  T  V  H
  801 AGACGCAGGT GTTGAAACAG ACGCAGGTAG TACTATGTCT ACTGTCCACG
      TCTGCGTCCA CAACTTTGTC TGCGTCCATC ATGATACAGA TGACAGGTGC

      +1 E  I  L  C  K  L  S  L  E  G  V  H  S
  851 AAATCCTGTG CAAGCTCAGC TTGGAGGGTG TTCATTCTAC ACCCCCAAGT
      TTTAGGACAC GTTCGAGTCG AACCTCCCAC AAGTAAGATG TGGGGGTTCA

            BamHI
            ------
  901 GCCGGATCC
      CGGCCTAGG
```

FIGURE 44
SEQUENCE ID NO 21-22

Sequence and Translation: EYFP-DEVD-MAP4-EBFP

This sequence codes for a bi-functional caspase-3/cytoskeleton biosensor. The chimeric protein consists (in order) of EYFP fluorescent protein (italic), a KGDEVDG caspase recognition site (bold text), the full-length MAP4 cDNA (unformatted text), and a C-terminal EBFP fluorescent protein (underlined).

```
  +1  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D
   1  ATGGTGAGCAAG GGCGAGGAGCTG TTCACCGGGGTG GTGCCCATCCTG GTCGAGCTGGAC
      TACCACTCGTTC CCGCTCCTCGAC AAGTGGCCCCAC CACGGGTAGGAC CAGCTCGACCTG

  +1  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
  61  GGCGACGTAAAC GGCCACAAGTTC AGCGTGTCCGGC GAGGGCGAGGGC GATGCCACCTAC
      CCGCTGCATTTG CCGGTGTTCAAG TCGCACAGGCCG CTCCCGCTCCCG CTACGGTGGATG

  +1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T
 121  GGCAAGCTGACC CTGAAGTTCATC TGCACCACCGGC AAGCTGCCCGTG CCCTGGCCCACC
      CCGTTCGACTGG GACTTCAAGTAG ACGTGGTGGCCG TTCGACGGGCAC GGGACCGGGTGG

  +1  L  V  T  T  F  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K
 181  CTCGTGACCACC TTCGGCTACGGC CTGCAGTGCTTC GCCCGCTACCCC GACCACATGAAG
      GAGCACTGGTGG AAGCCGATGCCG GACGTCACGAAG CGGGCGATGGGG CTGGTGTACTTC

  +1  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
 241  CAGCACGACTTC TTCAAGTCCGCC ATGCCCGAAGGC TACGTCCAGGAG CGCACCATCTTC
      GTCGTGCTGAAG AAGTTCAGGCGG TACGGGCTTCCG ATGCAGGTCCTC GCGTGGTAGAAG

  +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L
 301  TTCAAGGACGAC GGCAACTACAAG ACCCGCGCCGAG GTGAAGTTCGAG GGCGACACCCTG
      AAGTTCCTGCTG CCGTTGATGTTC TGGGCGCGGCTC CACTTCAAGCTC CCGCTGTGGGAC

  +1  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
 361  GTGAACCGCATC GAGCTGAAGGGC ATCGACTTCAAG GAGGACGGCAAC ATCCTGGGGCAC
      CACTTGGCGTAG CTCGACTTCCCG TAGCTGAAGTTC CTCCTGCCGTTG TAGGACCCCGTG
```

```
  +1  G  I  K  V   N  F  K  I   R  H  N  I   E  D  G  S   V  Q  L  A
481  GGCATCAAGGTG AACTTCAAGATC CGCCACAACATC GAGGACGGCAGC GTGCAGCTCGCC
     CCGTAGTTCCAC TTGAAGTTCTAG GCGGTGTTGTAG CTCCTGCCGTCG CACGTCGAGCGG

  +1  D  H  Y  Q   Q  N  T  P   I  G  D  G   P  V  L  L   P  D  N  H
541  GACCACTACCAG CAGAACACCCCC ATCGGCGACGGC CCCGTGCTGCTG CCCGACAACCAC
     CTGGTGATGGTC GTCTTGTGGGGG TAGCCGCTGCCG GGGCACGACGAC GGGCTGTTGGTG

  +1  Y  L  S  Y   Q  S  A  L   S  K  D  P   N  E  K  R   D  H  M  V
601  TACCTGAGCTAC CAGTCCGCCCTG AGCAAAGACCCC AACGAGAAGCGC GATCACATGGTC
     ATGGACTCGATG GTCAGGCGGGAC TCGTTTCTGGGG TTGCTCTTCGCG CTAGTGTACCAG

  +1  L  L  E  F   V  T  A  A   G  I  T  L   G  M  D  E   L  Y  K  K
661  CTGCTGGAGTTC GTGACCGCCGCC GGGATCACTCTC GGCATGGACGAG CTGTACAAGAAG
     GACGACCTCAAG CACTGGCGGCGG CCCTAGTGAGAG CCGTACCTGCTC GACATGTTCTTC

  +1  G  D  E  V   D  G  M  A   D  L  S  L   V  D  A  L   T  E  P  P
                                                HincII
                                                --------
721  GGAGACGAAGTG GACGGAATGGCC GACCTCAGTCTT GTGGATGCGTTG ACAGAACCACCT
     CCTCTGCTTCAC CTGCCTTACCGG CTGGAGTCAGAA CACCTACGCAAC TGTCTTGGTGGA

  +1  P  E  I  E   G  E  I  K   R  D  F  M   A  A  L  E   A  E  P  Y
781  CCAGAAATTGAG GGAGAAATAAAG CGAGACTTCATG GCTGCGCTGGAG GCAGAGCCCTAT
     GGTCTTTAACTC CCTCTTTATTTC GCTCTGAAGTAC CGACGCGACCTC CGTCTCGGGATA

  +1  D  D  I  V   G  E  T  V   E  K  T  E   F  I  P  L   L  D  G  D
841  GATGACATCGTG GGAGAAACTGTG GAGAAAACTGAG TTTATTCCTCTC CTGGATGGTGAT
     CTACTGTAGCAC CCTCTTTGACAC CTCTTTTGACTC AAATAAGGAGAG GACCTACCACTA

  +1  E  K  T  G   N  S  E  S   K  K  K  P   C  L  D  T   S  Q  V  E
901  GAGAAAACCGGG AACTCAGAGTCC AAAAAGAAACCC TGCTTAGACACT AGCCAGGTTGAA
     CTCTTTTGGCCC TTGAGTCTCAGG TTTTTCTTTGGG ACGAATCTGTGA TCGGTCCAACTT

  +1  G  I  P  S   S  K  P  T   L  L  A  N   G  D  H  G   M  E  G  N
961  GGTATCCCATCT TCTAAACCAACA CTCCTAGCCAAT GGTGATCATGGA ATGGAGGGGAAT
     CCATAGGGTAGA AGATTTGGTTGT GAGGATCGGTTA CCACTAGTACCT TACCTCCCCTTA

  +1  N  T  A  G   S  P  T  D   F  L  E  E   R  V  D  Y   P  D  Y  Q
1021 AACACTGCAGGG TCTCCAACTGAC TTCCTTGAAGAG AGAGTGGACTAT CCGGATTATCAG
     TTGTGACGTCCC AGAGGTTGACTG AAGGAACTTCTC TCTCACCTGATA GGCCTAATAGTC

  +1  S  S  Q  N   W  P  E  D   A  S  F  C   F  Q  P  Q   Q  V  L  D
                                HindIII
                                ------
1081 AGCAGCCAGAAC TGGCCAGAAGAT GCAAGCTTTTGT TTCCAGCCTCAG CAAGTGTTAGAT
     TCGTCGGTCTTG ACCGGTCTTCTA CGTTCGAAAACA AAGGTCGGAGTC GTTCACAATCTA
```

```
   +1  V  S  S  G   P  T  N  A   S  A  F  T   E  R  D  N   P  S  E  D
1201  GTCTCCAGTGGA CCCACGAACGCT TCTGCATTTACA GAGCGAGACAAT CCTTCAGAAGAC
      CAGAGGTCACCT GGGTGCTTGCGA AGACGTAAATGT CTCGCTCTGTTA GGAAGTCTTCTG

   +1  S  Y  G  M   L  P  C  D   S  F  A  S   T  A  V  V   S  Q  E  W
1261  AGTTACGGTATG CTTCCCTGTGAC TCATTTGCTTCC ACGGCTGTTGTA TCTCAGGAGTGG
      TCAATGCCATAC GAAGGGACACTG AGTAAACGAAGG TGCCGACAACAT AGAGTCCTCACC

   +1  S  V  G  A   P  N  S  P   C  S  E  S   C  V  S  P   E  V  T  I
1321  TCTGTGGGAGCC CCAAACTCTCCA TGTTCAGAGTCC TGTGTCTCCCCA GAGGTTACTATA
      AGACACCCTCGG GGTTTGAGAGGT ACAAGTCTCAGG ACACAGAGGGGT CTCCAATGATAT

   +1  E  T  L  Q   P  A  T  E   L  S  K  A   A  E  V  E   S  V  K  E
1381  GAAACCCTACAG CCAGCAACAGAG CTCTCCAAGGCA GCAGAAGTGGAA TCAGTGAAAGAG
      CTTTGGGATGTC GGTCGTTGTCTC GAGAGGTTCCGT CGTCTTCACCTT AGTCACTTTCTC

   +1  Q  L  P  A   K  A  L  E   T  M  A  E   Q  T  T  D   V  V  H  S
                                                    BstXI
                                                 -------------
                                                              ApaLI
                                                              ------
1441  CAGCTGCCAGCT AAAGCATTGGAA ACGATGGCAGAG CAGACCACTGAT GTGGTGCACTCT
      GTCGACGGTCGA TTTCGTAACCTT TGCTACCGTCTC GTCTGGTGACTA CACCACGTGAGA

   +1  P  S  T  D   T  T  P  G   P  D  T  E   A  A  L  A   K  D  I  E
1501  CCATCCACAGAC ACAACACCAGGC CCAGACACAGAG GCAGCACTGGCT AAAGACATAGAA
      GGTAGGTGTCTG TGTTGTGGTCCG GGTCTGTGTCTC CGTCGTGACCGA TTTCTGTATCTT

   +1  E  I  T  K   P  D  V  I   L  A  N  V   T  Q  P  S   T  E  S  D
1561  GAGATCACCAAG CCAGATGTGATA TTGGCAAATGTC ACGCAGCCATCT ACTGAATCGGAT
      CTCTAGTGGTTC GGTCTACACTAT AACCGTTTACAG TGCGTCGGTAGA TGACTTAGCCTA

   +1  M  F  L  A   Q  D  M  E   L  L  T  G   T  E  A  A   H  A  N  N
1621  ATGTTCCTGGCC CAGGACATGGAA CTACTCACAGGA ACAGAGGCAGCC CACGCTAACAAT
      TACAAGGACCGG GTCCTGTACCTT GATGAGTGTCCT TGTCTCCGTCGG GTGCGATTGTTA

   +1  I  I  L  P   T  E  P  D   E  S  S  T   K  D  V  A   P  P  M  E
1681  ATCATATTGCCT ACAGAACCAGAC GAATCTTCAACC AAGGATGTAGCA CCACCTATGGAA
      TAGTATAACGGA TGTCTTGGTCTG CTTAGAAGTTGG TTCCTACATCGT GGTGGATACCTT

   +1  E  E  I  V   P  G  N  D   T  T  S  P   K  E  T  E   T  T  L  P
1741  GAAGAAATTGTC CCAGGCAATGAT ACGACATCCCCC AAAGAAACAGAG ACAACACTTCCA
      CTTCTTTAACAG GGTCCGTTACTA TGCTGTAGGGGG TTTCTTTGTCTC TGTTGTGAAGGT

   +1  I  K  M  D   L  A  P  P   E  D  V  L   L  T  K  E   T  E  L  A
1801  ATAAAAATGGAC TTGGCACCACCT GAGGATGTGTTA CTTACCAAAGAA ACAGAACTAGCC
      TATTTTTACCTG AACCGTGGTGGA CTCCTACACAAT GAATGGTTTCTT TGTCTTGATCGG
```

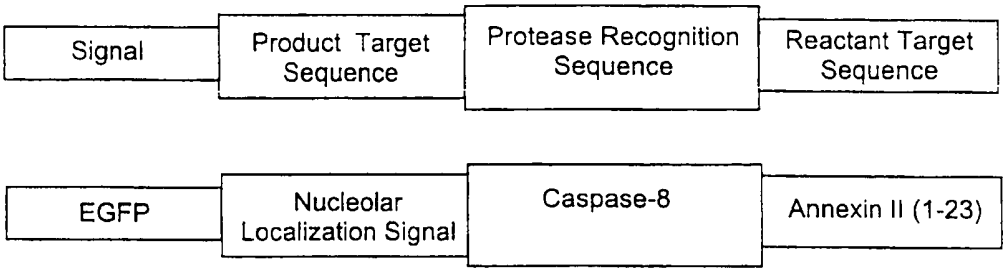
```
   +1  R  S  A  E  I  P  V  A  Q  E  T  V  V  S  E  T  E  V  V  L
1921  CGCTCTGCAGAA ATACCTGTGGCT CAGGAGACAGTG GTCTCAGAAACA GAGGTGGTCCTG
      GCGAGACGTCTT TATGGACACCGA GTCCTCTGTCAC CAGAGTCTTTGT CTCCACCAGGAC

   +1  A  T  E  V  V  L  P  S  D  P  I  T  T  L  T  K  D  V  T  L
1981  GCAACAGAAGTG GTACTGCCCTCA GATCCCATAACA ACATTGACAAAG GATGTGACACTC
      CGTTGTCTTCAC CATGACGGGAGT CTAGGGTATTGT TGTAACTGTTTC CTACACTGTGAG

   +1  P  L  E  A  E  R  P  L  V  T  D  M  T  P  S  L  E  T  E  M
2041  CCCTTAGAAGCA GAGAGACCGTTG GTGACGGACATG ACTCCATCTCTG GAAACAGAAATG
      GGGAATCTTCGT CTCTCTGGCAAC CACTGCCTGTAC TGAGGTAGAGAC CTTTGTCTTTAC

   +1  T  L  G  K  E  T  A  P  P  T  E  T  N  L  G  M  A  K  D  M
                                          ApoI
                                         -------
2101  ACCCTAGGCAAA GAGACAGCTCCA CCCACAGAAACA AATTTGGGCATG GCCAAAGACATG
      TGGGATCCGTTT CTCTGTCGAGGT GGGTGTCTTTGT TTAAACCCGTAC CGGTTTCTGTAC

   +1  S  P  L  P  E  S  E  V  T  L  G  K  D  V  V  I  L  P  E  T
2161  TCTCCACTCCCA GAATCAGAAGTG ACTCTGGGCAAG GACGTGGTTATA CTTCCAGAAACA
      AGAGGTGAGGGT CTTAGTCTTCAC TGAGACCCGTTC CTGCACCAATAT GAAGGTCTTTGT

   +1  K  V  A  E  F  N  N  V  T  P  L  S  E  E  E  V  T  S  V  K
2221  AAGGTGGCTGAG TTTAACAATGTG ACTCCACTTTCA GAAGAAGAGGTA ACCTCAGTCAAG
      TTCCACCGACTC AAATTGTTACAC TGAGGTGAAAGT CTTCTTCTCCAT TGGAGTCAGTTC

   +1  D  M  S  P  S  A  E  T  E  A  P  L  A  K  N  A  D  L  H  S
2281  GACATGTCTCCG TCTGCAGAAACA GAGGCTCCCCTG GCTAAGAATGCT GATCTGCACTCA
      CTGTACAGAGGC AGACGTCTTTGT CTCCGAGGGGAC CGATTCTTACGA CTAGACGTGAGT

   +1  G  T  E  L  I  V  D  N  S  M  A  P  A  S  D  L  A  L  P  L
2341  GGAACAGAGCTG ATTGTGGACAAC AGCATGGCTCCA GCCTCCGATCTT GCACTGCCCTTG
      CCTTGTCTCGAC TAACACCTGTTG TCGTACCGAGGT CGGAGGCTAGAA CGTGACGGGAAC

   +1  E  T  K  V  A  T  V  P  I  K  D  K  G  T  V  Q  T  E  E  K
2401  GAAACAAAAGTA GCAACAGTTCCA ATTAAAGACAAA GGAACTGTACAG ACTGAAGAAAAA
      CTTTGTTTTCAT CGTTGTCAAGGT TAATTTCTGTTT CCTTGACATGTC TGACTTCTTTTT

   +1  P  R  E  D  S  Q  L  A  S  M  Q  H  K  G  Q  S  T  V  P  P
                                                HincII
                                               -------
2461  CCACGTGAAGAC TCCCAGTTAGCA TCTATGCAGCAC AAGGGACAGTCA ACAGTACCTCCT
      GGTGCACTTCTG AGGGTCAATCGT AGATACGTCGTG TTCCCTGTCAGT TGTCATGGAGGA

   +1  C  T  A  S  P  E  P  V  K  A  A  E  Q  M  S  T  L  P  I  D
                                                   AccI
                                                  ------
2521  TGCACGGCTTCA CCAGAACCAGTC AAAGCTGCAGAA CAAATGTCTACC TTACCAATAGAT
```

```
   +1  P  C  S  G   V  S  R  Q   E  E  A  K   A  A  V  G   V  T  G  N
2641  CCTTGCTCAGGA GTATCCCGGCAA GAAGAAGCAAAG GCTGCTGTAGGT GTGACTGGAAAT
      GGAACGAGTCCT CATAGGGCCGTT CTTCTTCGTTTC CGACGACATCCA CACTGACCTTTA

   +1  D  I  T  T   P  P  N  K   E  P  P  P   S  P  E  K   K  A  K  P
2701  GACATCACTACC CCGCCAAACAAG GAGCCACCACCA AGCCCAGAAAAG AAAGCAAAGCCT
      CTGTAGTGATGG GGCGGTTTGTTC CTCGGTGGTGGT TCGGGTCTTTTC TTTCGTTTCGGA

   +1  L  A  T  T   Q  P  A  K   T  S  T  S   K  A  K  T   Q  P  T  S
2761  TTGGCCACCACT CAACCTGCAAAG ACTTCAACATCG AAAGCCAAAACA CAGCCCACTTCT
      AACCGGTGGTGA GTTGGACGTTTC TGAAGTTGTAGC TTTCGGTTTTGT GTCGGGTGAAGA

   +1  L  P  K  Q   P  A  P  T   T  S  G  G   L  N  K  K   P  M  S  L
2821  CTCCCTAAGCAA CCAGCTCCCACC ACCTCTGGTGGG TTGAATAAAAAA CCCATGAGCCTC
      GAGGGATTCGTT GGTCGAGGGTGG TGGAGACCACCC AACTTATTTTTT GGGTACTCGGAG

   +1  A  S  G  S   V  P  A  A   P  H  K  R   P  A  A  A   T  A  T  A
2881  GCCTCAGGCTCA GTGCCAGCTGCC CCACACAAACGC CCTGCTGCTGCC ACTGCTACTGCC
      CGGAGTCCGAGT CACGGTCGACGG GGTGTGTTTGCG GGACGACGACGG TGACGATGACGG

   +1  R  P  S  T   L  P  A  R   D  V  K  P   K  P  I  T   E  A  K  V
2941  AGGCCTTCCACC CTACCTGCCAGA GACGTGAAGCCA AAGCCAATTACA GAAGCTAAGGTT
      TCCGGAAGGTGG GATGGACGGTCT CTGCACTTCGGT TTCGGTTAATGT CTTCGATTCCAA

   +1  A  E  K  R   T  S  P  S   K  P  S  S   A  P  A  L   K  P  G  P
3001  GCCGAAAAGCGG ACCTCTCCATCC AAGCCTTCATCT GCCCCAGCCCTC AAACCTGGACCT
      CGGCTTTTCGCC TGGAGAGGTAGG TTCGGAAGTAGA CGGGGTCGGGAG TTTGGACCTGGA

   +1  K  T  T  P   T  V  S  K   A  T  S  P   S  T  L  V   S  T  G  P
3061  AAAACCACCCCA ACCGTTTCAAAA GCCACATCTCCC TCAACTCTTGTT TCCACTGGACCA
      TTTTGGTGGGGT TGGCAAAGTTTT CGGTGTAGAGGG AGTTGAGAACAA AGGTGACCTGGT

   +1  S  S  R  S   P  A  T  T   L  P  K  R   P  T  S  I   K  T  E  G
                                    BglI
                                 ------------
3121  AGTAGTAGAAGT CCAGCTACAACT CTGCCTAAGAGG CCAACCAGCATC AAGACTGAGGGG
      TCATCATCTTCA GGTCGATGTTGA GACGGATTCTCC GGTTGGTCGTAG TTCTGACTCCCC

   +1  K  P  A  D   V  K  R  M   T  A  K  S   A  S  A  D   L  S  R  S
3181  AAACCTGCTGAT GTCAAAAGGATG ACTGCTAAGTCT GCCTCAGCTGAC TTGAGTCGCTCA
      TTTGGACGACTA CAGTTTTCCTAC TGACGATTCAGA CGGAGTCGACTG AACTCAGCGAGT

   +1  K  T  T  S   A  S  S  V   K  R  N  T   T  P  T  G   A  A  P  P
3241  AAGACCACCTCT GCCAGTTCTGTG AAGAGAAACACC ACTCCCACTGGG GCAGCACCCCCA
      TTCTGGTGGAGA CGGTCAAGACAC TTCTCTTTGTGG TGAGGGTGACCC CGTCGTGGGGGT

   +1  A  G  M  T   S  T  R  V   K  P  M  S   A  P  S  R   S  S  G  A
                      XhoI
```

```
   +1  L  S  V  D   K  K  P  T   S  T  K  P   S  S  S  A   P  R  V  S
3361  CTTTCTGTGGAC AAGAAGCCCACT TCCACTAAGCCT AGCTCCTCTGCT CCCAGGGTGAGC
      GAAAGACACCTG TTCTTCGGGTGA AGGTGATTCGGA TCGAGGAGACGA GGGTCCCACTCG

   +1  R  L  A  T   T  V  S  A   P  D  L  K   S  V  R  S   K  V  G  S
3421  CGCCTGGCCACA ACTGTTTCTGCC CCTGACCTGAAG AGTGTTCGCTCC AAGGTCGGCTCT
      GCGGACCGGTGT TGACAAAGACGG GGACTGGACTTC TCACAAGCGAGG TTCCAGCCGAGA

   +1  T  E  N  I   K  H  Q  P   G  G  G  R   A  K  V  E   K  K  T  E
3481  ACAGAAAACATC AAACACCAGCCT GGAGGAGGCCGG GCCAAAGTAGAG AAAAAAACAGAG
      TGTCTTTTGTAG TTTGTGGTCGGA CCTCCTCCGGCC CGGTTTCATCTC TTTTTTTGTCTC

   +1  A  A  T  T   A  G  K  P   E  P  N  A   V  T  K  A   A  G  S  I
3541  GCAGCTACCACA GCTGGGAAGCCT GAACCTAATGCA GTCACTAAAGCA GCCGGCTCCATT
      CGTCGATGGTGT CGACCCTTCGGA CTTGGATTACGT CAGTGATTTCGT CGGCCGAGGTAA

   +1  A  S  A  Q   K  P  P  A   G  K  V  Q   I  V  S  K   K  V  S  Y
          ApaLI
          ~~~~~~
3601  GCGAGTGCACAG AAACCGCCTGCT GGGAAAGTCCAG ATAGTATCCAAA AAAGTGAGCTAC
      CGCTCACGTGTC TTTGGCGGACGA CCCTTTCAGGTC TATCATAGGTTT TTTCACTCGATG

   +1  S  H  I  Q   S  K  C  V   S  K  D  N   I  K  H  V   P  G  C  G
3661  AGTCATATTCAA TCCAAGTGTGTT TCCAAGGACAAT ATTAAGCATGTC CCTGGATGTGGC
      TCAGTATAAGTT AGGTTCACACAA AGGTTCCTGTTA TAATTCGTACAG GGACCTACACCG

   +1  N  V  Q  I   Q  N  K  K   V  D  I  S   K  V  S  S   K  C  G  S
3721  AATGTTCAGATT CAGAACAAGAAA GTGGACATATCC AAGGTCTCCTCC AAGTGTGGGTCC
      TTACAAGTCTAA GTCTTGTTCTTT CACCTGTATAGG TTCCAGAGGAGG TTCACACCCAGG

   +1  K  A  N  I   K  H  K  P   G  G  G  D   V  K  I  E   S  Q  K  L
3781  AAAGCTAATATC AAGCACAAGCCT GGTGGAGGAGAT GTCAAGATTGAA AGTCAGAAGTTG
      TTTCGATTATAG TTCGTGTTCGGA CCACCTCCTCTA CAGTTCTAACTT TCAGTCTTCAAC

   +1  N  F  K  E   K  A  Q  A   K  V  G  S   L  D  N  V   G  H  F  P
                                      BamHI
                                      ~~~~~~
3841  AACTTCAAGGAG AAGGCCCAAGCC AAAGTGGGATCC CTTGATAACGTT GGCCACTTTCCT
      TTGAAGTTCCTC TTCCGGGTTCGG TTTCACCCTAGG GAACTATTGCAA CCGGTGAAAGGA

   +1  A  G  G  A   V  K  T  E   G  G  G  S   E  A  L  P   C  P  G  P
3901  GCAGGAGGTGCC GTGAAGACTGAG GGCGGTGGCAGT GAGGCCCTTCCG TGTCCAGGCCCC
      CGTCCTCCACGG CACTTCTGACTC CCGCCACCGTCA CTCCGGGAAGGC ACAGGTCCGGGG

   +1  P  A  G  E   E  P  V  I   P  E  A  A   P  D  R  G   A  P  T  S
3961  CCCGCTGGGGAG GAGCCAGTCATC CCTGAGGCTGCG CCTGACCGTGGC GCCCCTACTTCA
      GGGCGACCCCTC CTCGGTCAGTAG GGACTCCGACGC GGACTGGCACCG CGGGGATGAAGT
```

```
   +1  T  L  D  S  Q  I  Q  E  T  S  I  M  V  S  K  G  E  E  L  F
4081  ACCTTGGACAGC CAGATCCAGGAG ACAAGCATCATG GTGAGCAAGGGC GAGGAGCTGTTC
      TGGAACCTGTCG GTCTAGGTCCTC TGTTCGTAGTAC CACTCGTTCCCG CTCCTCGACAAG

   +1  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S
4141  ACCGGGGTGGTG CCCATCCTGGTC GAGCTGGACGGC GACGTAAACGGC CACAAGTTCAGC
      TGGCCCCACCAC GGGTAGGACCAG CTCGACCTGCCG CTGCATTTGCCG GTGTTCAAGTCG

   +1  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C
4201  GTGTCCGGCGAG GGCGAGGGCGAT GCCACCTACGGC AAGCTGACCCTG AAGTTCATCTGC
      CACAGGCCGCTC CCGCTCCCGCTA CGGTGGATGCCG TTCGACTGGGAC TTCAAGTAGACG

   +1  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  H  G  V
4261  ACCACCGGCAAG CTGCCCGTGCCC TGGCCCACCCTC GTGACCACCCTG ACCCACGGCGTG
      TGGTGGCCGTTC GACGGGCACGGG ACCGGGTGGGAG CACTGGTGGGAC TGGGTGCCGCAC

   +1  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M
4321  CAGTGCTTCAGC CGCTACCCCGAC CACATGAAGCAG CACGACTTCTTC AAGTCCGCCATG
      GTCACGAAGTCG GCGATGGGGCTG GTGTACTTCGTC GTGCTGAAGAAG TTCAGGCGGTAC

   +1  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T
4381  CCCGAAGGCTAC GTCCAGGAGCGC ACCATCTTCTTC AAGGACGACGGC AACTACAAGACC
      GGGCTTCCGATG CAGGTCCTCGCG TGGTAGAAGAAG TTCCTGCTGCCG TTGATGTTCTGG

   +1  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I
4441  CGCGCCGAGGTG AAGTTCGAGGGC GACACCCTGGTG AACCGCATCGAG CTGAAGGGCATC
      GCGCGGCTCCAC TTCAAGCTCCCG CTGTGGGACCAC TTGGCGTAGCTC GACTTCCCGTAG

   +1  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  F  N  S  H
4501  GACTTCAAGGAG GACGGCAACATC CTGGGGCACAAG CTGGAGTACAAC TTCAACAGCCAC
      CTGAAGTTCCTC CTGCCGTTGTAG GACCCCGTGTTC GACCTCATGTTG AAGTTGTCGGTG

   +1  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R
4561  AACGTCTATATC ATGGCCGACAAG CAGAAGAACGGC ATCAAGGTGAAC TTCAAGATCCGC
      TTGCAGATATAG TACCGGCTGTTC GTCTTCTTGCCG TAGTTCCACTTG AAGTTCTAGGCG

   +1  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I
4621  CACAACATCGAG GACGGCAGCGTG CAGCTCGCCGAC CACTACCAGCAG AACACCCCCATC
      GTGTTGTAGCTC CTGCCGTCGCAC GTCGAGCGGCTG GTGATGGTCGTC TTGTGGGGGTAG

   +1  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S
4681  GGCGACGGCCCC GTGCTGCTGCCC GACAACCACTAC CTGAGCACCCAG TCCGCCCTGAGC
      CCGCTGCCGGGG CACGACGACGGG CTGTTGGTGATG GACTCGTGGGTC AGGCGGGACTCG

   +1  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G
4741  AAAGACCCCAAC GAGAAGCGCGAT CACATGGTCCTG CTGGAGTTCGTG ACCGCCGCCGGG
      TTTCTGGGGTTG CTCTTCGCGCTA GTGTACCAGGAC GACCTCAAGCAC TGGCGGCGGCCC
```

Figure 45


Signal
Product Target Sequence
Protease Recognition Sequence
Reactant Target Sequence
EGFP
Nucleolar Localization Signal
Caspase-8
Annexin II (1-23)

SEQ ID 23-24

FIGURE 46

Caspase 8 with nucleolus sequence: This molecule consists of GFP (underline), nucleolus sequence is in italic , caspase sequence is in bold and annexin II is double underline.

```
      +1 M A S K G E E L F T G V V P I L V
       1 ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
         TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1 E L D G D V N G H K F S V S G E
      51 TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
         ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1 G E G D A T Y G K L T L K F I C T
     101 GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
         CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1 T G K L P V P W P T L V T T L C Y
                    NcoI
                    -------
     151 ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
         TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1 G V Q C F S R Y P D H M K R H D
     201 TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
         ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1 F F K S A M P E G Y V Q E R T I F
     251 TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
         AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1 F K D D G N Y K T R A E V K F E G
     301 TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
         AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1 D T L V N R I E L K G I D F K E
     351 TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
         ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1 D G N I L G H K L E Y N Y N S H N
     401 ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
         TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1 V Y I M A D K Q K N G I K V N F K
     451 GTATACATCA TGGCAGACAA ACAAAAGAAT GGAATCAAAG TGAACTTCAA

         [illegible]
         CTGGGCGGTG TTGTAACTTC TACCTTCGCA AGTTGATCGT CTGGTAATAG

      +1 Q Q N T P I G D G P V L L P D N H
     551 AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
```

```
          TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA

      +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
     601  TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
          ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

      +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
     651  CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
          GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

      +1 M  D  E  L  Y  N  S  G  R  K  R  I  R  T  Y  L  K
     701 TGGATGAACT GTACAACTCC GGAAGAAAAC GTATACGTAC TTACCTCAAG
         ACCTACTTGA CATGTTGAGG CCTTCTTTTG CATATGCATG AATGGAGTTC

      +1 S  C  R  R  M  K  R  S  G  F  E  M  S  R  P  I  P
           PstI
         -------
     751 TCCTGCAGGC GGATGAAAAG AAGTGGTTTT GAGATGTCTC GACCTATTCC
         AGGACGTCCG CCTACTTTTC TTCACCAAAA CTCTACAGAG CTGGATAAGG

      +1  S  H  L  T  R  S  A  G  V  E  T  D  A  G  V  E
     801 TTCCCACCTT ACTCGATCGG CAGGTGTTGA AACAGACGCA GGTGTTGAAA
         AAGGGTGGAA TGAGCTAGCC GTCCACAACT TTGTCTGCGT CCACAACTTT

      +1 T  D  A  G  V  E  T  D  A  G  V  E  T  D  A  G  S
     851 CAGACGCAGG TGTTGAAACA GACGCAGGTG TTGAAACAGA CGCAGGTAGT
         GTCTGCGTCC ACAACTTTGT CTGCGTCCAC AACTTTGTCT GCGTCCATCA

      +1 T  M  S  T  V  H  E  I  L  C  K  L  S  L  E  G  V
     901 ACTATGTCTA CTGTCCACGA AATCCTGTGC AAGCTCAGCT TGGAGGGTGT
         TGATACAGAT GACAGGTGCT TTAGGACACG TTCGAGTCGA ACCTCCCACA

      +1  H  S  T  P  P  S  A  G  S
                                BamHI
                               ------
     951 TCATTCTACA CCCCCAAGTG CCGGATCC
         AGTAAGATGT GGGGGTTCAC GGCCTAGG
```

SEQ ID 25-26

FIGURE 47

Caspase 3- substrate with nucleolus sequence : This molecule consists of GFP(underline) , nucleolus sequence is in italic, caspase sequence is in bold and annexin II is double underline.

```
      +1  M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1  ATGGCTAGCA AAGGAGAAGA ACTCTTCACT GGAGTTGTCC CAATTCTTGT
          TACCGATCGT TTCCTCTTCT TGAGAAGTGA CCTCAACAGG GTTAAGAACA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51  TGAATTAGAT GGTGATGTTA ACGGCCACAA GTTCTCTGTC AGTGGAGAGG
          ACTTAATCTA CCACTACAAT TGCCGGTGTT CAAGAGACAG TCACCTCTCC

      +1 G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101  GTGAAGGTGA TGCAACATAC GGAAAACTTA CCCTGAAGTT CATCTGCACT
          CACTTCCACT ACGTTGTATG CCTTTTGAAT GGGACTTCAA GTAGACGTGA

      +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  C  Y
                              NcoI
                             -------
     151  ACTGGCAAAC TGCCTGTTCC ATGGCCAACA CTAGTCACTA CTCTGTGCTA
          TGACCGTTTG ACGGACAAGG TACCGGTTGT GATCAGTGAT GAGACACGAT

      +1  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D
     201  TGGTGTTCAA TGCTTTTCAA GATACCCGGA TCATATGAAA CGGCATGACT
          ACCACAAGTT ACGAAAAGTT CTATGGGCCT AGTATACTTT GCCGTACTGA

      +1 F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251  TTTTCAAGAG TGCCATGCCC GAAGGTTATG TACAGGAAAG GACCATCTTC
          AAAAGTTCTC ACGGTACGGG CTTCCAATAC ATGTCCTTTC CTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301  TTCAAAGATG ACGGCAACTA CAAGACACGT GCTGAAGTCA AGTTTGAAGG
          AAGTTTCTAC TGCCGTTGAT GTTCTGTGCA CGACTTCAGT TCAAACTTCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351  TGATACCCTT GTTAATAGAA TCGAGTTAAA AGGTATTGAC TTCAAGGAAG
          ACTATGGGAA CAATTATCTT AGCTCAATTT TCCATAACTG AAGTTCCTTC

      +1 D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401  ATGGCAACAT TCTGGGACAC AAATTGGAAT ACAACTATAA CTCACACAAT
          TACCGTTGTA AGACCCTGTG TTTAACCTTA TGTTGATATT GAGTGTGTTA

      +1  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
```

```
 501  GACCCGCCAC AACATTGAAG ATGGAAGCGT TCAACTAGCA GACCATTATC
      CTGGGCGGTG TTGTAACTTC TACCTTCGCA AGTTGATCGT CTGGTAATAG

 +1  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
 551  AACAAAATAC TCCAATTGGC GATGGCCCTG TCCTTTTACC AGACAACCAT
      TTGTTTTATG AGGTTAACCG CTACCGGGAC AGGAAAATGG TCTGTTGGTA

 +1  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D
 601  TACCTGTCCA CACAATCTGC CCTTTCGAAA GATCCCAACG AAAAGAGAGA
      ATGGACAGGT GTGTTAGACG GGAAAGCTTT CTAGGGTTGC TTTTCTCTCT

 +1   H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
 651  CCACATGGTC CTTCTTGAGT TTGTAACAGC TGCTGGGATT ACACATGGCA
      GGTGTACCAG GAAGAACTCA AACATTGTCG ACGACCCTAA TGTGTACCGT

 +1  M  D  E  L  Y  N  S  G  R  K  R  I  R  T  Y  L  K
 701  TGGATGAACT GTACAACTCC GGAAGAAAAC GTATACGTAC TTACCTCAAG
      ACCTACTTGA CATGTTGAGG CCTTCTTTTG CATATGCATG AATGGAGTTC

 +1  S  C  R  R  M  K  R  S  G  F  E  M  S  R  P  I  P
       PstI
      ------
 751  TCCTGCAGGC GGATGAAAAG AAGTGGTTTT GAGATGTCTC GACCTATTCC
      AGGACGTCCG CCTACTTTTC TTCACCAAAA CTCTACAGAG CTGGATAAGG

 +1   S  H  L  T  R  S  Y  E  K  G  I  P  V  E  T  D
 801  TTCCCACCTT ACTCGATCGT ATGAAAAAGG AATACCAGTT GAAACAGACA
      AAGGGTGGAA TGAGCTAGCA TACTTTTTCC TTATGGTCAA CTTTGTCTGT

 +1  S  E  E  Q  A  Y  S  T  M  S  T  V  H  E  I  L  C
            HindIII
            -------
 851  GCGAAGAGCA AGCTTATAGT ACTATGTCTA CTGTCCACGA AATCCTGTGC
      CGCTTCTCGT TCGAATATCA TGATACAGAT GACAGGTGCT TTAGGACACG

 +1  K  L  S  L  E  G  V  H  S  T  P  P  S  A  G  S
                                                BamHI
                                               ------
 901  AAGCTCAGCT TGGAGGGTGT TCATTCTACA CCCCCAAGTG CCGGATCC
      TTCGAGTCGA ACCTCCCACA AGTAAGATGT GGGGGTTCAC GGCCTAGG
```

FIGURE 48

SEQ ID 27-28
Sequence: NLS-FRED25-cellubrevin

```
       +1  M  R  R  K   R  Q  K  A   S  K  G  E   E  L  F  T   G  V  V
P

                                   NheI
                                   -------

       1  ATGAGAAGAAAA CGACAAAAGGCT AGCAAAGGAGAA GAACTCTTCACT
GGAGTTGTCCCA
          TACTCTTCTTTT GCTGTTTTCCGA TCGTTTCCTCTT CTTGAGAAGTGA CCTCAACAGGGT

       +1  I  L  V  E   L  D  G  D   V  N  G  H   K  F  S  V   S  G  E
G

                                   HincII
                                   ------
      61  ATTCTTGTTGAA TTAGATGGTGAT GTTAACGGCCAC AAGTTCTCTGTC
AGTGGAGAGGGT
          TAAGAACAACTT AATCTACCACTA CAATTGCCGGTG TTCAAGAGACAG TCACCTCTCCCA

       +1  E  G  D  A   T  Y  G  K   L  T  L  K   F  I  C  T   T  G  K
L

     121  GAAGGTGATGCA ACATACGGAAAA CTTACCCTGAAG TTCATCTGCACT
ACTGGCAAACTG
          CTTCCACTACGT TGTATGCCTTTT GAATGGGACTTC AAGTAGACGTGA TGACCGTTTGAC

       +1  P  V  P  W   P  T  L  V   T  T  L  C   Y  G  V  Q   C  F  S
R

     181  CCTGTTCCATGG CCAACACTAGTC ACTACTCTGTGC TATGGTGTTCAA
TGCTTTTCAAGA
          GGACAAGGTACC GGTTGTGATCAG TGATGAGACACG ATACCACAAGTT ACGAAAAGTTCT

       +1  Y  P  D  H   M  K  R  H   D  F  F  K   S  A  M  P   E  G  Y
V

                    NdeI
                    -------
     241  TACCCGGATCAT ATGAAACGGCAT GACTTTTTCAAG AGTGCCATGCCC
GAAGGTTATGTA
          ATGGGCCTAGTA TACTTTGCCGTA CTGAAAAAGTTC TCACGGTACGGG CTTCCAATACAT

                AVaI
                -----
     301  CAGGAAAGGACC ATCTTCTTCAAA GATGACGGCAAC TACAAGACACGT
GCTGAAGTCAAG
```

```
      +1  F  E  G  D   T  L  V  N   R  I  E  L   K  G  I  D   F  K  E
D

    361  TTTGAAGGTGAT ACCCTTGTTAAT AGAATCGAGTTA AAAGGTATTGAC
TTCAAGGAAGAT
         AAACTTCCACTA TGGGAACAATTA TCTTAGCTCAAT TTTCCATAACTG AAGTTCCTTCTA

      +1  G  N  I  L   G  H  K  L   E  Y  N  Y   N  S  H  N   V  Y  I
M
                                                              AccI
                                                             -------

    421  GGCAACATTCTG GGACACAAATTG GAATACAACTAT AACTCACACAAT
GTATACATCATG
         CCGTTGTAAGAC CCTGTGTTTAAC CTTATGTTGATA TTGAGTGTGTTA CATATGTAGTAC

      +1  A  D  K  Q   K  N  G  I   K  V  N  F   K  T  R  H   N  I  E
D

    481  GCAGACAAACAA AAGAATGGAATC AAAGTGAACTTC AAGACCCGCCAC
AACATTGAAGAT
         CGTCTGTTTGTT TTCTTACCTTAG TTTCACTTGAAG TTCTGGGCGGTG TTGTAACTTCTA

      +1  G  S  V  Q   L  A  D  H   Y  Q  Q  N   T  P  I  G   D  G  P
V

    541  GGAAGCGTTCAA CTAGCAGACCAT TATCAACAAAAT ACTCCAATTGGC
GATGGCCCTGTC
         CCTTCGCAAGTT GATCGTCTGGTA ATAGTTGTTTTA TGAGGTTAACCG CTACCGGGACAG

      +1  L  L  P  D   N  H  Y  L   S  T  Q  S   A  L  S  K   D  P  N
E
                                                               BstYI
                                                              --------

    601  CTTTTACCAGAC AACCATTACCTG TCCACACAATCT GCCCTTTCGAAA
GATCCCAACGAA
         GAAAATGGTCTG TTGGTAATGGAC AGGTGTGTTAGA CGGGAAAGCTTT CTAGGGTTGCTT

      +1  K  R  D  H   M  V  L  L   E  F  V  T   A  A  G  I   T  H  G
M
                        AvaII
                        ------

      +1  D  E  L  Y   N  T  G  M   N  T  G  V   P  S  G  S   G  A  A
T

                        AgeI      AccI
```

```
                          ------  -------

     721  GATGAACTGTAC AACACCGGTATG TCTACAGGTGTG CCTTCGGGGTCA A
GTGCTGCCACT
          CTACTTGACATG TTGTGGCCATAC AGATGTCCACAC GGAAGCCCCAGT
TCACGACGGTGA

       +1  G  S  N  R   R  L  Q  Q   T  Q  N  Q   V  D  E  V   V  D  I
M
                                                              HincII
                                                              ------
     781  GGCAGTAATCGA AGACTCCAGCAG ACACAAAATCAA GTAGATGAGGTG
GTTGACATCATG
          CCGTCATTAGCT TCTGAGGTCGTC TGTGTTTTAGTT CATCTACTCCAC
CAACTGTAGTAC

       +1  R  V  N  V   D  K  V  L   E  R  D  Q   K  L  S  E   L  D  D
R
     841  AGAGTCAATGTG GATAAGGTGTTA GAAAGAGACCAG AAGCTCTCGGAG
CTAGATGACCGC
          TCTCAGTTACAC CTATTCCACAAT CTTTCTCTGGTC TTCGAGAGCCTC
GATCTACTGGCG

       +1  A  D  A  L   Q  A  G  A   S  Q  F  E   T  S  A  A   K  L  K
R
                    PstI       BanI
                   --------   -------
     901  GCAGATGCACTG CAGGCAGGTGCC TCGCAGTTTGAA ACAAGTGCTGCC
AAGTTGAAGAGA
          CGTCTACGTGAC GTCCGTCCACGG AGCGTCAAACTT TGTTCACGACGG
TTCAACTTCTCT

       +1  K  Y  W  W   K  N  C  K   M  W  A  I   G  I  S  V   L  V  I
I
                                                            EcoRII
                                                             ------
     961  AAGTATTGGTGG AAGAACTGCAAG ATGTGGGCGATA GGGATCAGTGTC
CTGGTGATCATT
          TTCATAACCACC TTCTTGACGTTC TACACCCGCTAT CCCTAGTCACAG
GACCACTAGTAA

       +1  V  I  I  I   I  V  W  C   V  S  *
    1021  GTCATCATCATC ATCGTGTGGTGT GTCTCTTAA
          CAGTAGTAGTAG TAGCACACCACA CAGAGAATT

atgagaagaaaacgacaaaaggctagcaaaggagaagaactcttcactggagttgtcccaattcttgttgaatta
gatggtgatgttaacggccacaagttctctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacc
```

gtggttgacatcatgagagtcaatgtggataaggtgttagaaagagaccagaagctctcggagctagatgaccgc
gcagatgcactgcaggcaggtgcctcgcagtttgaaacaagtgctgccaagttgaagagaaagtattggtggaag
aactgcaagatgtgggcgatagggatcagtgtcctggtgatcattgtcatcatcatcgtgtggtgtgtctct
taa

FIGURE 49

SEQ ID 29-30
Sequence: NLS-FRED25-synaptobrevin

```
       +1  M  R  R  K   R  Q  K  A   S  K  G  E   E  L  F  T   G  V  V
P
                                  NheI
                                  -------
       1  ATGAGAAGAAAA CGACAAAAGGCT AGCAAAGGAGAA GAACTCTTCACT
GGAGTTGTCCCA
          TACTCTTCTTTT GCTGTTTTCCGA TCGTTTCCTCTT CTTGAGAAGTGA
CCTCAACAGGGT

       +1  I  L  V  E   L  D  G  D   V  N  G  H   K  F  S  V   S  G  E
G
                                     HincII
                                     ------
      61  ATTCTTGTTGAA TTAGATGGTGAT GTTAACGGCCAC AAGTTCTCTGTC
AGTGGAGAGGGT
          TAAGAACAACTT AATCTACCACTA CAATTGCCGGTG TTCAAGAGACAG
TCACCTCTCCCA

       +1  E  G  D  A   T  Y  G  K   L  T  L  K   F  I  C  T   T  G  K
L
     121  GAAGGTGATGCA ACATACGGAAAA CTTACCCTGAAG TTCATCTGCACT
ACTGGCAAACTG
          CTTCCACTACGT TGTATGCCTTTT GAATGGGACTTC AAGTAGACGTGA
TGACCGTTTGAC

       +1  P  V  P  W   P  T  L  V   T  T  L  C   Y  G  V  Q   C  F  S
R
     181  CCTGTTCCATGG CCAACACTAGTC ACTACTCTGTGC TATGGTGTTCAA
TGCTTTTCAAGA
          GGACAAGGTACC GGTTGTGATCAG TGATGAGACACG ATACCACAAGTT
ACGAAAAGTTCT

       +1  Y  P  D  H   M  K  R  H   D  F  F  K   S  A  M  P   E  G  Y
V
                    NdeI
                    -------
     241  TACCCGGATCAT ATGAAACGGCAT GACTTTTTCAAG AGTGCCATGCCC
GAAGGTTATGTA
          ATGGGCCTAGTA TACTTTGCCGTA CTGAAAAAGTTC TCACGGTACGGG
CTTCCAATACAT

       +1  Q  E  R  T   I  F  F  K   D  D  G  N   Y  K  T  R   A  E  V
K
```

```
       +1  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
D
     361  TTTGAAGGTGAT ACCCTTGTTAAT AGAATCGAGTTA AAAGGTATTGAC
 TTCAAGGAAGAT
          AAACTTCCACTA TGGGAACAATTA TCTTAGCTCAAT TTTCCATAACTG
AAGTTCCTTCTA

       +1  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I
M
                                                          AccI
                                                         ------
     421  GGCAACATTCTG GGACACAAATTG GAATACAACTAT AACTCACACAAT
GTATACATCATG
          CCGTTGTAAGAC CCTGTGTTTAAC CTTATGTTGATA TTGAGTGTGTTA
CATATGTAGTAC

       +1  A  D  K  Q  K  N  G  I  K  V  N  F  K  T  R  H  N  I  E
D
     481  GCAGACAAACAA AAGAATGGAATC AAAGTGAACTTC AAGACCCGCCAC
AACATTGAAGAT
          CGTCTGTTTGTT TTCTTACCTTAG TTTCACTTGAAG TTCTGGGCGGTG
TTGTAACTTCTA

       +1  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P
V
     541  GGAAGCGTTCAA CTAGCAGACCAT TATCAACAAAAT ACTCCAATTGGC
GATGGCCCTGTC
          CCTTCGCAAGTT GATCGTCTGGTA ATAGTTGTTTTA TGAGGTTAACCG
CTACCGGGACAG

       +1  L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S  K  D  P  N
E
                                                       BstYI
                                                      -------
     601  CTTTTACCAGAC AACCATTACCTG TCCACACAATCT GCCCTTTCGAAA
GATCCCAACGAA
          GAAAATGGTCTG TTGGTAATGGAC AGGTGTGTTAGA CGGGAAAGCTTT
CTAGGGTTGCTT

       +1  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G
M
     661  AAGAGAGACCAC ATGGTCCTTCTT GAGTTTGTAACA GCTGCTGGGATT ACACATGGCATG
          TTCTCTCTGGTG TACCAGGAAGAA CTCAAACATTGT CGACGACCCTAA TGTGTACCGTAC

       +1  D  E  L  Y  N  T  G  M  S  T  G  P  T  A  A  T  G  S  N
T

          CTACTTGACATG TTGTGGCCATAC AGATGTCCAGGT TGACGACGGTGA CCGTCATTAGCT
```

```
      +1  R  L  Q  Q   T  Q  N  Q   V  D  E  V   V  D  I  M   R  V  N
V
                                                                HincII
                                                                ------
     781  AGACTTCAGCAG ACACAAAATCAA GTAGATGAGGTG GTGGACATAATG
CGAGTTAACGTG
          TCTGAAGTCGTC TGTGTTTTAGTT CATCTACTCCAC CACCTGTATTAC
GCTCAATTGCAC

      +1  D  K  V  L   E  R  D  Q   K  L  S  E   L  D  D  R   A  D  A
L
                                                                   PstI
                                                                   ---
     841  GACAAGGTTCTG GAAAGAGACCAG AAGCTCTCTGAG TTAGACGACCGT
GCAGACGCACTG
          CTGTTCCAAGAC CTTTCTCTGGTC TTCGAGAGACTC AATCTGCTGGCA
CGTCTGCGTGAC

      +1  Q  A  G  A   S  Q  F  E   T  S  A  A   K  L  K  R   K  Y  W
W
          PstI
          ---
     901  CAGGCAGGCGCT TCTCAATTTGAA ACGAGCGCAGCC AAGTTGAAGAGG
AAATATTGGTGG
          GTCCGTCCGCGA AGAGTTAAACTT TGCTCGCGTCGG TTCAACTTCTCC
TTTATAACCACC

      +1  K  N  C  K   M  W  A  I   G  I  T  V   L  V  I  F   I  I  I
I
     961  AAGAATTGCAAG ATGTGGGCAATC GGGATTACTGTT CTGGTTATCTTC
ATCATCATCATC
          TTCTTAACGTTC TACACCCGTTAG CCCTAATGACAA GACCAATAGAAG
TAGTAGTAGTAG

      +1  I  V  W  V   V  S  S  *
    1021  ATCGTGTGGGTT GTCTCTTCATGA
          TAGCACACCCAA CAGAGAAGTACT
```

```
atgagaagaaaacgacaaaaggctagcaaaggagaagaactcttcactggagttgtcccaattcttgttgaatta
gatggtgatgttaacggccacaagttctctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacc
ctgaagttcatctgcactactggcaaactgcctgttccatggccaacactagtcactactctgtgctatggtgtt
caatgcttttcaagatacccggatcatatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgta
caggaaaggaccatcttcttcaaagatgacggcaactacaagacacgtgctgaagtcaagtttgaaggtgatacc
cttgttaatagaatcgagttaaaaggtattgacttcaaggaagatggcaacattctgggacacaaattggaatac
aactataactcacacaatgtatacatcatggcagacaaacaaaagaatggaatcaaagtgaacttcaagacccgc
```

ggtccaactgctgccactggcagtaatcgaagacttcagcagacacaaaatcaagtagatgaggtggtggacata
atgcgagttaacgtggacaaggttctggaaagagaccagaagctctctgagttagacgaccgtgcagacgcactg
caggcaggcgcttctcaatttgaaacgagcgcagccaagttgaagaggaaatattggtggaagaattgcaagatg
tgggcaatcgggattactgttctggttatcttcatcatcatcatcatcgtgtgggttgtctcttcatga

Figure 50

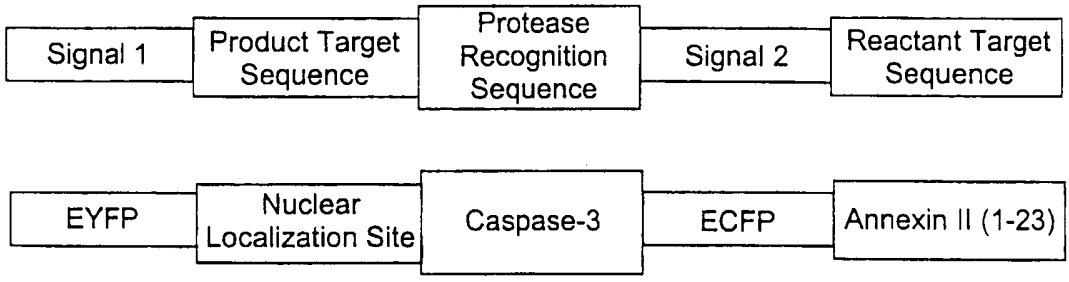

Signal 1
Product Target Sequence
Protease Recognition Sequence
Signal 2
Reactant Target Sequence
EYFP
Nuclear Localization Site
Caspase-3
ECFP
Annexin II (1-23)

FIGURE 51

SEQUENCE ID NO 31-32

Sequence and Translation: NLS-EYFP-DEVD-MAPKDM-EBFP

This sequence codes for a ratiometric caspase biosensor. The chimeric protein consists (in order) of an NLS (double underline), EYFP fluorescent protein (italic), a KGDEVDG caspase recognition site (bold text), the projection domain of MAP4 for size exclusion from the nucleus (unformatted text), and a C-terminal EBFP fluorescent protein (underlined).

```
+1   M  R  P  R  R  K  V  S  K  G  E  E  L  F  T  G  V  V  P  I
1    ATGAGGCCCAGA AGAAAGGTGAGC AAGGGCGAGGAG CTGTTCACCGGG GTGGTGCCCATC
     TACTCCGGGTCT TCTTTCCACTCG TTCCCGCTCCTC GACAAGTGGCCC CACCACGGGTAG

+1   L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E
61   CTGGTCGAGCTG GACGGCGACGTA AACGGCCACAAG TTCAGCGTGTCC GGCGAGGGCGAG
     GACCAGCTCGAC CTGCCGCTGCAT TTGCCGGTGTTC AAGTCGCACAGG CCGCTCCCGCTC

+1   G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P
121  GGCGATGCCACC TACGGCAAGCTG ACCCTGAAGTTC ATCTGCACCACC GGCAAGCTGCCC
     CCGCTACGGTGG ATGCCGTTCGAC TGGGACTTCAAG TAGACGTGGTGG CCGTTCGACGGG

+1   V  P  W  P  T  L  V  T  T  F  G  Y  G  L  Q  C  F  A  R  Y
181  GTGCCCTGGCCC ACCCTCGTGACC ACCTTCGGCTAC GGCCTGCAGTGC TTCGCCCGCTAC
     CACGGGACCGGG TGGGAGCACTGG TGGAAGCCGATG CCGGACGTCACG AAGCGGGCGATG

+1   P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q
241  CCCGACCACATG AAGCAGCACGAC TTCTTCAAGTCC GCCATGCCCGAA GGCTACGTCCAG
     GGGCTGGTGTAC TTCGTCGTGCTG AAGAAGTTCAGG CGGTACGGGCTT CCGATGCAGGTC

+1   E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F
301  GAGCGCACCATC TTCTTCAAGGAC GACGGCAACTAC AAGACCCGCGCC GAGGTGAAGTTC
     CTCGCGTGGTAG AAGAAGTTCCTG CTGCCGTTGATG TTCTGGGCGCGG CTCCACTTCAAG

+1   E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G
361  GAGGGCGACACC CTGGTGAACCGC ATCGAGCTGAAG GGCATCGACTTC AAGGAGGACGGC
     CTCCCGCTGTGG GACCACTTGGCG TAGCTCGACTTC CCGTAGCTGAAG TTCCTCCTGCCG

+1   N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A
```

```
     CTGTTCGTCTTC TTGCCGTAGTTC CACTTGAAGTTC TAGGCGGTGTTG TAGCTCCTGCCG

  +1 S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L
 541 AGCGTGCAGCTC GCCGACCACTAC CAGCAGAACACC CCCATCGGCGAC GGCCCCGTGCTG
     TCGCACGTCGAG CGGCTGGTGATG GTCGTCTTGTGG GGGTAGCCGCTG CCGGGGCACGAC

  +1 L  P  D  N  H  Y  L  S  Y  Q  S  A  L  S  K  D  P  N  E  K
 601 CTGCCCGACAAC CACTACCTGAGC TACCAGTCCGCC CTGAGCAAAGAC CCCAACGAGAAG
     GACGGGCTGTTG GTGATGGACTCG ATGGTCAGGCGG GACTCGTTTCTG GGGTTGCTCTTC

  +1 R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D
 661 CGCGATCACATG GTCCTGCTGGAG TTCGTGACCGCC GCCGGGATCACT CTCGGCATGGAC
     GCGCTAGTGTAC CAGGACGACCTC AAGCACTGGCGG CGGCCCTAGTGA GAGCCGTACCTG

  +1 E  L  Y  K  K  G  D  E  V  D  G  A  D  L  S  L  V  D  A  L
                                                      HincII
                                                      ----
 721 GAGCTGTACAAG AAGGGAGACGAA GTGGACGGAGCC GACCTCAGTCTT GTGGATGCGTTG
     CTCGACATGTTC TTCCCTCTGCTT CACCTGCCTCGG CTGGAGTCAGAA CACCTACGCAAC

  +1 T  E  P  P  P  E  I  E  G  E  I  K  R  D  F  M  A  A  L  E
     HincII
     --
 781 ACAGAACCACCT CCAGAAATTGAG GGAGAAATAAAG CGAGACTTCATG GCTGCGCTGGAG
     TGTCTTGGTGGA GGTCTTTAACTC CCTCTTTATTTC GCTCTGAAGTAC CGACGCGACCTC

  +1 A  E  P  Y  D  D  I  V  G  E  T  V  E  K  T  E  F  I  P  L
 841 GCAGAGCCCTAT GATGACATCGTG GGAGAAACTGTG GAGAAAACTGAG TTTATTCCTCTC
     CGTCTCGGGATA CTACTGTAGCAC CCTCTTTGACAC CTCTTTTGACTC AAATAAGGAGAG

  +1 L  D  G  D  E  K  T  G  N  S  E  S  K  K  K  P  C  L  D  T
 901 CTGGATGGTGAT GAGAAAACCGGG AACTCAGAGTCC AAAAAGAAACCC TGCTTAGACACT
     GACCTACCACTA CTCTTTTGGCCC TTGAGTCTCAGG TTTTTCTTTGGG ACGAATCTGTGA

  +1 S  Q  V  E  G  I  P  S  S  K  P  T  L  L  A  N  G  D  H  G
 961 AGCCAGGTTGAA GGTATCCCATCT TCTAAACCAACA CTCCTAGCCAAT GGTGATCATGGA
     TCGGTCCAACTT CCATAGGGTAGA AGATTTGGTTGT GAGGATCGGTTA CCACTAGTACCT

  +1 M  E  G  N  N  T  A  G  S  P  T  D  F  L  E  E  R  V  D  Y
1021 ATGGAGGGGAAT AACACTGCAGGG TCTCCAACTGAC TTCCTTGAAGAG AGAGTGGACTAT
     TACCTCCCCTTA TTGTGACGTCCC AGAGGTTGACTG AAGGAACTTCTC TCTCACCTGATA

  +1 P  D  Y  Q  S  S  Q  N  W  P  E  D  A  S  F  C  F  Q  P  Q
                                          HindIII
                                          ------
1081 CCGGATTATCAG AGCAGCCAGAAC TGGCCAGAAGAT GCAAGCTTTTGT TTCCAGCCTCAG
     GGCCTAATAGTC TCGTCGGTCTTG ACCGGTCTTCTA CGTTCGAAAACA AAGGTCGGAGTC

  +1 Q  V  L  D  T  D  Q  A  E  P  F  N  E  H  R  D  D  G  L  A
```

```
   +1  D  L  L  F   V  S  S  G   P  T  N  A   S  A  F  T   E  R  D  N
       BglII
       -----
1201   GATCTGCTCTTT GTCTCCAGTGGA CCCACGAACGCT TCTGCATTTACA GAGCGAGACAAT
       CTAGACGAGAAA CAGAGGTCACCT GGGTGCTTGCGA AGACGTAAATGT CTCGCTCTGTTA

   +1  P  S  E  D   S  Y  G  M   L  P  C  D   S  F  A  S   T  A  V  V
1261   CCTTCAGAAGAC AGTTACGGTATG CTTCCCTGTGAC TCATTTGCTTCC ACGGCTGTTGTA
       GGAAGTCTTCTG TCAATGCCATAC GAAGGGACACTG AGTAAACGAAGG TGCCGACAACAT

   +1  S  Q  E  W   S  V  G  A   P  N  S  P   C  S  E  S   C  V  S  P
1321   TCTCAGGAGTGG TCTGTGGGAGCC CCAAACTCTCCA TGTTCAGAGTCC TGTGTCTCCCCA
       AGAGTCCTCACC AGACACCCTCGG GGTTTGAGAGGT ACAAGTCTCAGG ACACAGAGGGGT

   +1  E  V  T  I   E  T  L  Q   P  A  T  E   L  S  K  A   A  E  V  E
1381   GAGGTTACTATA GAAACCCTACAG CCAGCAACAGAG CTCTCCAAGGCA GCAGAAGTGGAA
       CTCCAATGATAT CTTTGGGATGTC GGTCGTTGTCTC GAGAGGTTCCGT CGTCTTCACCTT

   +1  S  V  K  E   Q  L  P  A   K  A  L  E   T  M  A  E   Q  T  T  D
                                                                BstXI
                                                               --------
1441   TCAGTGAAAGAG CAGCTGCCAGCT AAAGCATTGGAA ACGATGGCAGAG CAGACCACTGAT
       AGTCACTTTCTC GTCGACGGTCGA TTTCGTAACCTT TGCTACCGTCTC GTCTGGTGACTA

   +1  V  V  H  S   P  S  T  D   T  T  P  G   P  D  T  E   A  A  L  A
       BstXI
       ----
           ApaLI
           ------
1501   GTGGTGCACTCT CCATCCACAGAC ACAACACCAGGC CCAGACACAGAG GCAGCACTGGCT
       CACCACGTGAGA GGTAGGTGTCTG TGTTGTGGTCCG GGTCTGTGTCTC CGTCGTGACCGA

   +1  K  D  I  E   E  I  T  K   P  D  V  I   L  A  N  V   T  Q  P  S
1561   AAAGACATAGAA GAGATCACCAAG CCAGATGTGATA TTGGCAAATGTC ACGCAGCCATCT
       TTTCTGTATCTT CTCTAGTGGTTC GGTCTACACTAT AACCGTTTACAG TGCGTCGGTAGA

   +1  T  E  S  D   M  F  L  A   Q  D  M  E   L  L  T  G   T  E  A  A
1621   ACTGAATCGGAT ATGTTCCTGGCC CAGGACATGGAA CTACTCACAGGA ACAGAGGCAGCC
       TGACTTAGCCTA TACAAGGACCGG GTCCTGTACCTT GATGAGTGTCCT TGTCTCCGTCGG

   +1  H  A  N  N   I  I  L  P   T  E  P  D   E  S  S  T   K  D  V  A
1681   CACGCTAACAAT ATCATATTGCCT ACAGAACCAGAC GAATCTTCAACC AAGGATGTAGCA
       GTGCGATTGTTA TAGTATAACGGA TGTCTTGGTCTG CTTAGAAGTTGG TTCCTACATCGT

   +1  P  P  M  E   E  E  I  V   P  G  N  D   T  T  S  P   K  E  T  E
1741   CCACCTATGGAA GAAGAAATTGTC CCAGGCAATGAT ACGACATCCCCC AAAGAAACAGAG
       GGTGGATACCTT CTTCTTTAACAG GGTCCGTTACTA TGCTGTAGGGGG TTTCTTTGTCTC

   +1  T  T  L  P   I  K  M  D   L  A  P  P   E  D  V  L   L  T  K  E
```

BstXI

```
                          ~~~~~~~~~~~~~~
1861  ACAGAACTAGCC CCAGCCAAGGGC ATGGTTTCACTC TCAGAAATAGAA GAGGCTCTGGCA
      TGTCTTGATCGG GGTCGGTTCCCG TACCAAAGTGAG AGTCTTTATCTT CTCCGAGACCGT

  +1  K  N  D  V  R  S  A  E  I  P  V  A  Q  E  T  V  V  S  E  T
1921  AAGAATGATGTT CGCTCTGCAGAA ATACCTGTGGCT CAGGAGACAGTG GTCTCAGAAACA
      TTCTTACTACAA GCGAGACGTCTT TATGGACACCGA GTCCTCTGTCAC CAGAGTCTTTGT

  +1  E  V  V  L  A  T  E  V  V  L  P  S  D  P  I  T  T  L  T  K
1981  GAGGTGGTCCTG GCAACAGAAGTG GTACTGCCCTCA GATCCCATAACA ACATTGACAAAG
      CTCCACCAGGAC CGTTGTCTTCAC CATGACGGGAGT CTAGGGTATTGT TGTAACTGTTTC

  +1  D  V  T  L  P  L  E  A  E  R  P  L  V  T  D  M  T  P  S  L
2041  GATGTGACACTC CCCTTAGAAGCA GAGAGACCGTTG GTGACGGACATG ACTCCATCTCTG
      CTACACTGTGAG GGGAATCTTCGT CTCTCTGGCAAC CACTGCCTGTAC TGAGGTAGAGAC

  +1  E  T  E  M  T  L  G  K  E  T  A  P  P  T  E  T  N  L  G  M
                                                      ApoI
                                                    -------
2101  GAAACAGAAATG ACCCTAGGCAAA GAGACAGCTCCA CCCACAGAAACA AATTTGGGCATG
      CTTTGTCTTTAC TGGGATCCGTTT CTCTGTCGAGGT GGGTGTCTTTGT TTAAACCCGTAC

  +1  A  K  D  M  S  P  L  P  E  S  E  V  T  L  G  K  D  V  V  I
2161  GCCAAAGACATG TCTCCACTCCCA GAATCAGAAGTG ACTCTGGGCAAG GACGTGGTTATA
      CGGTTTCTGTAC AGAGGTGAGGGT CTTAGTCTTCAC TGAGACCCGTTC CTGCACCAATAT

  +1  L  P  E  T  K  V  A  E  F  N  N  V  T  P  L  S  E  E  E  V
2221  CTTCCAGAAACA AAGGTGGCTGAG TTTAACAATGTG ACTCCACTTTCA GAAGAAGAGGTA
      GAAGGTCTTTGT TTCCACCGACTC AAATTGTTACAC TGAGGTGAAAGT CTTCTTCTCCAT

  +1  T  S  V  K  D  M  S  P  S  A  E  T  E  A  P  L  A  K  N  A
2281  ACCTCAGTCAAG GACATGTCTCCG TCTGCAGAAACA GAGGCTCCCCTG GCTAAGAATGCT
      TGGAGTCAGTTC CTGTACAGAGGC AGACGTCTTTGT CTCCGAGGGGAC CGATTCTTACGA

  +1  D  L  H  S  G  T  E  L  I  V  D  N  S  M  A  P  A  S  D  L
2341  GATCTGCACTCA GGAACAGAGCTG ATTGTGGACAAC AGCATGGCTCCA GCCTCCGATCTT
      CTAGACGTGAGT CCTTGTCTCGAC TAACACCTGTTG TCGTACCGAGGT CGGAGGCTAGAA

  +1  A  L  P  L  E  T  K  V  A  T  V  P  I  K  D  K  G  M  V  S
2401  GCACTGCCCTTG GAAACAAAAGTA GCAACAGTTCCA ATTAAAGACAAA GGAATGGTGAGC
      CGTGACGGGAAC CTTTGTTTTCAT CGTTGTCAAGGT TAATTTCTGTTT CCTTACCACTCG

  +1  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V
2461  AAGGGCGAGGAG CTGTTCACCGGG GTGGTGCCCATC CTGGTCGAGCTG GACGGCGACGTA
      TTCCCGCTCCTC GACAAGTGGCCC CACCACGGGTAG GACCAGCTCGAC CTGCCGCTGCAT

  +1  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L
2521  AACGGCCACAAG TTCAGCGTGTCC GGCGAGGGCGAG GGCGATGCCACC TACGGCAAGCTG
      TTGCCGGTGTTC AAGTCGCACAGG CCGCTCCCGCTC CCGCTACGGTGG ATGCCGTTCGAC
```

```
      +1  T  L  T  H  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D
2641  ACCCTGACCCAC GGCGTGCAGTGC TTCAGCCGCTAC CCCGACCACATG AAGCAGCACGAC
      TGGGACTGGGTG CCGCACGTCACG AAGTCGGCGATG GGGCTGGTGTAC TTCGTCGTGCTG

      +1  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D
2701  TTCTTCAAGTCC GCCATGCCCGAA GGCTACGTCCAG GAGCGCACCATC TTCTTCAAGGAC
      AAGAAGTTCAGG CGGTACGGGCTT CCGATGCAGGTC CTCGCGTGGTAG AAGAAGTTCCTG

      +1  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R
2761  GACGGCAACTAC AAGACCCGCGCC GAGGTGAAGTTC GAGGGCGACACC CTGGTGAACCGC
      CTGCCGTTGATG TTCTGGGCGCGG CTCCACTTCAAG CTCCCGCTGTGG GACCACTTGGCG

      +1  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E
2821  ATCGAGCTGAAG GGCATCGACTTC AAGGAGGACGGC AACATCCTGGGG CACAAGCTGGAG
      TAGCTCGACTTC CCGTAGCTGAAG TTCCTCCTGCCG TTGTAGGACCCC GTGTTCGACCTC

      +1  Y  N  F  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K
2881  TACAACTTCAAC AGCCACAACGTC TATATCATGGCC GACAAGCAGAAG AACGGCATCAAG
      ATGTTGAAGTTG TCGGTGTTGCAG ATATAGTACCGG CTGTTCGTCTTC TTGCCGTAGTTC

      +1  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y
2941  GTGAACTTCAAG ATCCGCCACAAC ATCGAGGACGGC AGCGTGCAGCTC GCCGACCACTAC
      CACTTGAAGTTC TAGGCGGTGTTG TAGCTCCTGCCG TCGCACGTCGAG CGGCTGGTGATG

      +1  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S
3001  CAGCAGAACACC CCCATCGGCGAC GGCCCCGTGCTG CTGCCCGACAAC CACTACCTGAGC
      GTCGTCTTGTGG GGGTAGCCGCTG CCGGGGCACGAC GACGGGCTGTTG GTGATGGACTCG

      +1  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E
3061  ACCCAGTCCGCC CTGAGCAAAGAC CCCAACGAGAAG CGCGATCACATG GTCCTGCTGGAG
      TGGGTCAGGCGG GACTCGTTTCTG GGGTTGCTCTTC GCGCTAGTGTAC CAGGACGACCTC

      +1  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  *
3121  TTCGTGACCGCC GCCGGGATCACT CTCGGCATGGAC GAGCTGTACAAG TAG
      AAGCACTGGCGG CGGCCCTAGTGA GAGCCGTACCTG CTCGACATGTTC ATC
```

FIGURE 52
SEQ ID 33-34

This molecule consists of YFP (underline), NLS (italic), CP3 with multiple DEVD (bold), CFP (double underline) and Annexin II is dotted.

```
      +1  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V
       1  ATGGTGAGCA AGGGCGAGGA GCTGTTCACC GGGGTGGTGC CCATCCTGGT
          TACCACTCGT TCCCGCTCCT CGACAAGTGG CCCCACCACG GGTAGGACCA

      +1  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E
      51  CGAGCTGGAC GGCGACGTAA ACGGCCACAA GTTCAGCGTG TCCGGCGAGG
          GCTCGACCTG CCGCTGCATT TGCCGGTGTT CAAGTCGCAC AGGCCGCTCC

      +1  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T
     101  GCGAGGGCGA TGCCACCTAC GGCAAGCTGA CCCTGAAGTT CATCTGCACC
          CGCTCCCGCT ACGGTGGATG CCGTTCGACT GGGACTTCAA GTAGACGTGG

      +1  T  G  K  L  P  V  P  W  P  T  L  V  T  T  F  G  Y
     151  ACCGGCAAGC TGCCCGTGCC CTGGCCCACC CTCGTGACCA CCTTCGGCTA
          TGGCCGTTCG ACGGGCACGG GACCGGGTGG GAGCACTGGT GGAAGCCGAT

      +1  G  L  Q  C  F  A  R  Y  P  D  H  M  K  Q  H  D
                PstI
                ------
     201  CGGCCTGCAG TGCTTCGCCC GCTACCCCGA CCACATGAAG CAGCACGACT
          GCCGGACGTC ACGAAGCGGG CGATGGGGCT GGTGTACTTC GTCGTGCTGA

      +1  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F
     251  TCTTCAAGTC CGCCATGCCC GAAGGCTACG TCCAGGAGCG CACCATCTTC
          AGAAGTTCAG GCGGTACGGG CTTCCGATGC AGGTCCTCGC GTGGTAGAAG

      +1  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G
     301  TTCAAGGACG ACGGCAACTA CAAGACCCGC GCCGAGGTGA AGTTCGAGGG
          AAGTTCCTGC TGCCGTTGAT GTTCTGGGCG CGGCTCCACT TCAAGCTCCC

      +1  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E
     351  CGACACCCTG GTGAACCGCA TCGAGCTGAA GGGCATCGAC TTCAAGGAGG
          GCTGTGGGAC CACTTGGCGT AGCTCGACTT CCCGTAGCTG AAGTTCCTCC

      +1  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N
     401  ACGGCAACAT CCTGGGGCAC AAGCTGGAGT ACAACTACAA CAGCCACAAC
          TGCCGTTGTA GGACCCCGTG TTCGACCTCA TGTTGATGTT GTCGGTGTTG

      +1  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K
     451  GTCTATATCA TGGCCGACAA GCAGAAGAAC GGCATCAAGG TGAACTTCAA
```

```
        CTAGGCGGTG TTGTAGCTCC TGCCGTCGCA CGTCGAGCGG CTGGTGATGG
    +1  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H
551 AGCAGAACAC CCCCATCGGC GACGGCCCCG TGCTGCTGCC CGACAACCAC
    TCGTCTTGTG GGGGTAGCCG CTGCCGGGGC ACGACGACGG GCTGTTGGTG

    +1  Y  L  S  Y  Q  S  A  L  S  K  D  P  N  E  K  R  D
601 TACCTGAGCT ACCAGTCCGC CCTGAGCAAA GACCCCAACG AGAAGCGCGA
    ATGGACTCGA TGGTCAGGCG GGACTCGTTT CTGGGGTTGC TCTTCGCGCT

    +1  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L  G
651 TCACATGGTC CTGCTGGAGT TCGTGACCGC CGCCGGGATC ACTCTCGGCA
    AGTGTACCAG GACGACCTCA AGCACTGGCG GCGGCCCTAG TGAGAGCCGT

    +1  M  D  E  L  Y  K  S  G  R  R  K  R  Q  K  R  S  A
701 TGGACGAGCT GTACAAGTCC GGAAGAAGGA AACGACAAAA GCGATCGGCA
    ACCTGCTCGA CATGTTCAGG CCTTCTTCCT TTGCTGTTTT CGCTAGCCGT

    +1  G  D  E  V  D  A  G  D  E  V  D  A  G  D  E  V  D
751 GGTGACGAAG TTGATGCAGG TGACGAAGTT GATGCAGGTG ACGAAGTTGA
    CCACTGCTTC AACTACGTCC ACTGCTTCAA CTACGTCCAC TGCTTCAACT

    +1  A  G  D  E  V  D  A  G  S  T  M  V  S  K  G  E
801 TGCAGGTGAC GAAGTTGACG CAGGTAGTAC TATGGTGAGC AAGGGCGAGG
    ACGTCCACTG CTTCAACTGC GTCCATCATG ATACCACTCG TTCCCGCTCC

    +1  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V
851 AGCTGTTCAC CGGGGTGGTG CCCATCCTGG TCGAGCTGGA CGGCGACGTA
    TCGACAAGTG GCCCCACCAC GGGTAGGACC AGCTCGACCT GCCGCTGCAT

    +1  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y
901 AACGGCCACA AGTTCAGCGT GTCCGGCGAG GGCGAGGGCG ATGCCACCTA
    TTGCCGGTGT TCAAGTCGCA CAGGCCGCTC CCGCTCCCGC TACGGTGGAT

    +1  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V
951 CGGCAAGCTG ACCCTGAAGT TCATCTGCAC CACCGGCAAG CTGCCCGTGC
    GCCGTTCGAC TGGGACTTCA AGTAGACGTG GTGGCCGTTC GACGGGCACG

    +1  P  W  P  T  L  V  T  T  L  T  W  G  V  Q  C  F  S
1001 CCTGGCCCAC CCTCGTGACC ACCCTGACCT GGGGCGTGCA GTGCTTCAGC
    GGACCGGGTG GGAGCACTGG TGGGACTGGA CCCCGCACGT CACGAAGTCG

    +1  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P
1051 CGCTACCCCG ACCACATGAA GCAGCACGAC TTCTTCAAGT CCGCCATGCC
    GCGATGGGGC TGGTGTACTT CGTCGTGCTG AAGAAGTTCA GGCGGTACGG

    +1  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N
1101 CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC GACGGCAACT
    GCTTCCGATG CAGGTCCTCG CGTGGTAGAA GAAGTTCCTG CTGCCGTTGA

    +1  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H
```

```
1201  ATCGAGCTGA AGGGCATCGA CTTCAAGGAG GACGGCAACA TCCTGGGGCA
      TAGCTCGACT TCCCGTAGCT GAAGTTCCTC CTGCCGTTGT AGGACCCCGT

      +1   K   L   E   Y   N   Y   I   S   H   N   V   Y   I   T   A   D
1251  CAAGCTGGAG TACAACTACA TCAGCCACAA CGTCTATATC ACCGCCGACA
      GTTCGACCTC ATGTTGATGT AGTCGGTGTT GCAGATATAG TGGCGGCTGT

      +1 K   Q   K   N   G   I   K   A   N   F   K   I   R   H   N   I   E
1301  AGCAGAAGAA CGGCATCAAG GCCAACTTCA AGATCCGCCA CAACATCGAG
      TCGTCTTCTT GCCGTAGTTC CGGTTGAAGT TCTAGGCGGT GTTGTAGCTC

      +1   D   G   S   V   Q   L   A   D   H   Y   Q   Q   N   T   P   I   G
1351  GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA CCCCCATCGG
      CTGCCGTCGC ACGTCGAGCG GCTGGTGATG GTCGTCTTGT GGGGGTAGCC

      +1   D   G   P   V   L   L   P   D   N   H   Y   L   S   T   Q   S
1401  CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC ACCCAGTCCG
      GCTGCCGGGG CACGACGACG GGCTGTTGGT GATGGACTCG TGGGTCAGGC

      +1 A   L   S   K   D   P   N   E   K   R   D   H   M   V   L   L   E
1451  CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT CCTGCTGGAG
      GGGACTCGTT TCTGGGGTTG CTCTTCGCGC TAGTGTACCA GGACGACCTC

      +1 F   V   T   A   A   G   I   T   L   G   M   D   E   L   Y   K   M
1501  TTCGTGACCG CCGCCGGGAT CACTCTCGGC ATGGACGAGC TGTACAAGAT
      AAGCACTGGC GGCGGCCCTA GTGAGAGCCG TACCTGCTCG ACATGTTCTA

      +1   S   T   V   H   E   I   L   C   K   L   S   L   E   G   V   H
1551  GTCTACTGTC CACGAAATCC TGTGCAAGCT CAGCTTGGAG GGTGTTCATT
      CAGATGACAG GTGCTTTAGG ACACGTTCGA GTCGAACCTC CCACAAGTAA

      +1 S   T   P   P   S   A   G   S
                                BamHI
                                ------
1601  CTACACCCCC AAGTGCCGGA TCC
      GATGTGGGGG TTCACGGCCT AGG
```